Article

Sampling Importance Resampling Algorithm with Nonignorable Missing Response Variable Based on Smoothed Quantile Regression

Jingxuan Guo [1,†], Fuguo Liu [2,3,†], Wolfgang Karl Härdle [4,5,6], Xueliang Zhang [7], Kai Wang [7], Ting Zeng [7], Liping Yang [7] and Maozai Tian [7,*]

[1]   School of Statistics and Data Science, Beijing Wuzi University, Beijing 101149, China; guojingxuan@bwu.edu.cn
[2]   School of Statistics and Data Science, Xinjiang University of Finance and Economics, Urumqi 830012, China; lfg53880@cjc.edu.cn
[3]   School of Mathematics and Data Science, Changji University, Changji 831100, China
[4]   Department of Information Management and Finance, National Yang Ming Chiao Tung University, Taiwan 30010, China; haerdle@hu-berlin.de
[5]   Institute Digital Assets, Academy Economic Sciences, 010374 Bucharest, Romania
[6]   School of Business and Economics, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[7]   Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830011, China; shuxue2456@xjmu.edu.cn (X.Z.); kwang@xjmu.edu.cn (K.W.); zengting@xjmu.edu.cn (T.Z.); young_liping@xjmu.edu.cn (L.Y.)
[*]   Correspondence: mztian@ruc.edu.cn; Tel.: +86-135-2268-1878
[†]   The first two authors contributed equally to this work.

Abstract: The presence of nonignorable missing response variables often leads to complex conditional distribution patterns that cannot be effectively captured through mean regression. In contrast, quantile regression offers valuable insights into the conditional distribution. Consequently, this article places emphasis on the quantile regression approach to address nonrandom missing data. Taking inspiration from fractional imputation, this paper proposes a novel smoothed quantile regression estimation equation based on a sampling importance resampling (SIR) algorithm instead of nonparametric kernel regression methods. Additionally, we present an augmented inverse probability weighting (AIPW) smoothed quantile regression estimation equation to reduce the influence of potential misspecification in a working model. The consistency and asymptotic normality of the empirical likelihood estimators corresponding to the above estimating equations are proven under the assumption of a correctly specified parameter working model. Furthermore, we demonstrate that the AIPW estimation equation converges to an IPW estimation equation when a parameter working model is misspecified, thus illustrating the robustness of the AIPW estimation approach. Through numerical simulations, we examine the finite sample properties of the proposed method when the working models are both correctly specified and misspecified. Furthermore, we apply the proposed method to analyze HIV—CD4 data, thereby exploring variations in treatment effects and the influence of other covariates across different quantiles.

Keywords: empirical likelihood; nonignorable missing; quantile regression; sampling importance resampling

MSC: 62H25; 62F12

1. Introduction

Missing data analysis has gained significant attention in recent years. To analyze missing data, it is crucial to understand the response mechanism that leads to missing data. If the missingness of the variable of interest is conditionally independent of that variable,

the response mechanism is considered to be random or ignorable. Otherwise, the response mechanism is considered to be nonrandom or nonignorable. Dealing with nonrandom missing data presents greater challenges, which are evident in two aspects: Firstly, the assumed response model cannot be validated solely based on observed data; secondly, the model parameters may be unidentifiable.

To obtain meaningful inferences from incomplete data with nonrandom missingness, it is necessary to satisfy a set of identifying conditions [1,2]. Moreover, the accuracy of the methods based on parameter models is greatly influenced by the correct specification of the assumed parameter model [3]. Consequently, researchers aim to impose weaker model assumptions on the response mechanism to achieve robust results. The semiparametric response model was initially considered by Kim and Yu [4], but their proposed method necessitated a validation sample to estimate the model parameters. Similarly, Shao and Wang [5] examined the same semiparametric exponential tilting model and proposed a parameter estimation approach based on calibration estimation equations. A comprehensive review of parameter estimation methods for nonrandom missing data is provided by Kim and Shao [6].

Quantile regression, introduced by Koenker and Bassett [7], has become a widely used statistical analysis tool. It offers more adaptability and flexibility compared to mean regression. Notably, quantile regression does not require the assumption of error term distribution and demonstrates robustness against heavy-tailed errors and outliers. Furthermore, by considering regressions at different quantiles of the response variable, quantile regression enables the assessment of covariate effects at various quantiles and yields a more comprehensive understanding of the conditional distribution. However, there is a scarcity of literature on quantile regression for nonrandom missing data.

The nonsmoothness of the check function for standard quantile estimators makes it impossible to directly estimate the asymptotic covariance matrix [8]. As a result, the existing theoretical results for nonrandom missing mean regression cannot be directly extended to quantile regression.

The idea of smoothing nondifferentiable objective functions can be traced back to Horwitz [9], while Whang [10] introduced the smoothed empirical likelihood approach for quantile regression. Luo et al. [11] extended the aforementioned method to analyze data with random missingness; Zhang and Wang [12] further expanded it to handle cases of nonignorable missingness.

However, on the one hand, this method relies on the assumption of a parametric propensity missingness model, which introduces the risk of model misspecification. On the other hand, this method corrects estimation biases caused by missing data through inverse probability weighting but may not fully utilize the information from incomplete observations.

Regarding nonrandom missing data, previous studies have addressed the issue in different regression settings. Specifically, Niu et al. [13] and Bindele and Zhao [14] focused on estimation equation imputation in linear regression and rank regression, respectively. In the context of quantile regression, Chen et al. [15] introduced three missing quantile regression estimation equations: inverse probability weighting, estimation equation imputation, and an enhanced approach combining both methods. It is important to note that these studies assume a response mechanism with random missingness.

Moreover, the existing literature commonly utilizes kernel estimation methods [16] to estimate the conditional means involved in the imputation estimation equation. However, when the dimension of the covariates is high, the kernel estimation results can become unstable. To overcome the curse of dimensionality associated with multivariate nonparametric kernel estimation, Kim [17] proposed a parametric fractional imputation method for handling missing data. Additionally, Riddles et al. [18] extended this method to address the scenario of nonignorable missing data. They developed an EM algorithm based on a parameter working model derived from observed data and incorporated the parametric fractional imputation (FI) method. Nevertheless, these approaches heavily rely on parameter-based

response models, which renders them sensitive to model misspecification. Furthermore, the likelihood-based EM algorithm is not directly applicable to quantile regression.

Utilizing estimation equations, Paik and Larsen [19] incorporated a working model for observed data and employed a sampling importance resampling (SIR) algorithm to estimate the missing data and corresponding estimation equations. Building upon this, Wang et al. [20] and Song et al. [21] extended the logistic response model used in the aforementioned approaches to a develop a semiparametric exponential tilt model. However, in the absence of knowledge about the tilting parameter, these methods relied on validation samples.

In this paper, we propose a smoothed empirical likelihood approach for imputing quantile regression estimation equations with nonignorable missing data based on a semiparametric response model. The novel estimation equation guarantees the second-order differentiability of the objective function with respect to the parameter vector. The imputed values for the missing data were derived from a parameter working model and obtained using sampling importance resampling.

Although imputation estimation equations applying information from missing data compared to IPW estimation equations can enhance estimation efficiency, both theoretical and numerical experiments have shown that imputation estimation equations are sensitive to misspecification of the parameter working model. Therefore, to mitigate the impact of misspecification in the working model, this paper further proposes the AIPW smoothed quantile regression estimation equation. It is demonstrated that, when the working model is correctly specified, the asymptotic variance of the AIPW estimation equation shares the same form as the asymptotic variance of the nonparametric model estimator. Furthermore, even when the working model is misspecified, the estimates remain consistent.

The remaining sections of this paper are organized as follows. Section 2 establishes the semi-parametric response model and the AIPW quantile regression estimation equation, along with the algorithmic procedure for estimating the skewness parameter and quantile regression coefficients using importance resampling. Section 3 presents the large sample properties of the parameter estimators. Section 4 demonstrates the finite sample properties of the estimators through numerical simulations. Section 5 applies the proposed methodology to analyze the HIV—CD4 dataset.

2. Proposed Method

Consider a linear quantile regression model as follows:

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\theta}_\tau + \epsilon_i, \quad i = 1, \dots, n,$$

where Y_i is the response variable, $\mathbf{Z}_i$ is a fully observed q-dimensional covariate vector, $\boldsymbol{\theta}_\tau$ represents the unknown regression coefficient vector, ϵ_i denotes the random error term satisfying $P(\epsilon_i \leq 0 | \mathbf{Z}_i) = \tau$, $\tau \in (0,1)$, and the ϵ_i values are mutually independent. In the subsequent discussion, we will abbreviate $\boldsymbol{\theta}_\tau$ as $\boldsymbol{\theta}$.

If the response variables $Y_i, i = 1, \dots, n$ are fully observed, the quantile regression estimator of $\boldsymbol{\theta}$ is obtained by minimizing the following equation:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{Z}_i^\top \boldsymbol{\theta}), \tag{1}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, and $I(\cdot)$ is the indicator function. For a given τ, $\hat{\boldsymbol{\theta}}$ satisfies the following estimation equation:

$$\sum_{i=1}^{n} \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}) \approx 0, \tag{2}$$

where $\psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}) = \mathbf{Z}_i \psi_\tau(Y_i - \mathbf{Z}_i^\top \boldsymbol{\theta})$ when $Y_i - \mathbf{Z}_i^\top \boldsymbol{\theta} \neq 0$, and $\psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}) = 0$ otherwise. Here, $\psi_\tau(u) = I(u < 0) - \tau$.

In the scenario where the missingness of response variable Y_i is nonignorable, let δ_i denote the missing indicator. If Y_i is missing, $\delta_i = 1$; otherwise, $\delta_i = 0$. $(\boldsymbol{Z}_i, Y_i, \delta_i)$, $i = 1, \ldots, n$ represents an independent and identically distributed sample from $(\boldsymbol{Z}, Y, \delta)$. We establish a semiparametric exponential tilting model for missing propensity as follows:

$$P(\delta = 1 | \boldsymbol{Z}, Y) = \frac{1}{1 + \exp(-g(\boldsymbol{X}) + \gamma Y)}, \tag{3}$$

where $g(\cdot)$ is an unspecified function, $\boldsymbol{X} \subset \boldsymbol{Z}$ is a d-dimensional vector, and there exists an instrumental variable $\boldsymbol{V} = \boldsymbol{Z}/\boldsymbol{X}$ that is unrelated to δ given $(\boldsymbol{X}, Y)$.

Let $f_\kappa(Y, \boldsymbol{Z}|\boldsymbol{X})$, where $\kappa \in 0, 1$ denotes the conditional density of $(\boldsymbol{Z}, Y)$ of $\boldsymbol{X}$ when $\delta = \kappa$. Specifically, we have the following:

$$f_0(Y, \boldsymbol{Z}|\boldsymbol{X}, \gamma) = f_1(Y, \boldsymbol{Z}|\boldsymbol{X}) \times \frac{O(\boldsymbol{X}, Y)}{E\{O(\boldsymbol{X}, Y)|\boldsymbol{X}, \delta = 1\}}, \tag{4}$$

where $O(\boldsymbol{X}, Y) = \frac{P(\delta=0|\boldsymbol{X},Y)}{P(\delta=1|\boldsymbol{X},Y)} = \exp(-g(\boldsymbol{X}) + \gamma Y)$. For the quantile estimation equation $\psi(Y, \boldsymbol{Z}; \boldsymbol{\theta})$, let

$$\psi^{eei}(\boldsymbol{Z}, Y, \delta; \boldsymbol{\theta}, \gamma) = \delta \psi(Y, \boldsymbol{Z}; \boldsymbol{\theta}) + (1 - \delta)E(\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta})|\boldsymbol{X}, \delta = 0, \gamma), \tag{5}$$

it can be easily shown that $E(\psi(Y, \boldsymbol{Z}; \boldsymbol{\theta})|\boldsymbol{X}) = E(\psi^{eei}(Y, \boldsymbol{Z}; \boldsymbol{\theta})|\boldsymbol{X})$, where

$$E\{\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta})|\boldsymbol{X}, \delta = 0, \gamma\} = \frac{E\{\delta \exp(\gamma Y)\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta})|\boldsymbol{Z}\}}{E\{\delta \exp(\gamma Y)|\boldsymbol{Z}\}} := m_\psi^0(\boldsymbol{X}; \boldsymbol{\theta}, \gamma). \tag{6}$$

The nonparametric kernel estimate of Equation (6) is given by

$$\hat{m}_0(\boldsymbol{X}; \boldsymbol{\theta}, \gamma) = \sum_{i=1}^n \omega_i(\gamma)\psi(\boldsymbol{Z}_i, Y_i; \boldsymbol{\theta}),$$

where $\omega_i(\gamma) = \delta_i \exp(\gamma Y_i)\mathcal{K}_h(\boldsymbol{X} - \boldsymbol{X}_i)/\sum_{j=1}^n \delta_j \exp(\gamma Y_j)\mathcal{K}_h(\boldsymbol{X} - \boldsymbol{X}_j)$, $\mathcal{K}_h(u) = h^{-1}\mathcal{K}(u/h)$, and $\mathcal{K}(\cdot)$ is a d-dimensional kernel function with bandwidth h.

Due to the instability of the nonparametric multivariate kernel estimation of the aforementioned conditional expectation, this paper adopts Monte Carlo methods to estimate $m_\psi^0(\boldsymbol{X}; \boldsymbol{\theta}, \gamma)$. For simplicity of discussion, we consider the parameter assumption of the conditional distribution $f(Y|\boldsymbol{Z}, \delta = 1; \boldsymbol{\beta})$ of the observed response. This assumption can be verified easily using fully observed samples. Consequently, the conditional distribution of the response with nonrandom missingness satisfies

$$f_0(Y|\boldsymbol{Z}; \boldsymbol{\beta}, \gamma) = f_1(Y|\boldsymbol{Z}; \boldsymbol{\beta}) \times \frac{\exp(\gamma Y)}{E\{\exp(\gamma Y)|\boldsymbol{Z}, \delta = 1; \boldsymbol{\beta}\}}. \tag{7}$$

Let $Y_i^{(j)}, j = 1, \ldots, M$ be independent and identically distributed samples from $f(Y|\boldsymbol{Z}_i, \delta = 0; \boldsymbol{\beta}, \gamma)$. According to the law of large numbers, as $M \to \infty$, we have

$$\hat{m}_\psi^0(\boldsymbol{Z}_i; \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma) = \frac{1}{M}\sum_{j=1}^M \psi(\boldsymbol{Z}_i, Y_i^{*(j)}(\boldsymbol{\beta}, \gamma); \boldsymbol{\theta}) \xrightarrow{p} m_\psi^0(\boldsymbol{Z}_i; \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma).$$

To obtain a set of random realizations from $f_0(Y|\boldsymbol{Z}; \boldsymbol{\beta}, \gamma)$, the SIR algorithm [19] can be employed based on the parametric representation in (7) for a given $(\boldsymbol{\beta}, \gamma)$:

(1) Random samples $S_i = \{\tilde{Y}_i^{(k)}, k = 1, \ldots, M_2\}$ are drawn from $f(Y|\boldsymbol{Z}_i, \delta = 1; \boldsymbol{\beta})$.

(2) Calculate the adjustment weights for each sample point in S as

$$\omega_{ik}(\gamma) = \frac{\exp(\gamma \tilde{Y}_i^{(k)})}{\frac{1}{M_2} \sum_{j=1}^{M_2} \exp(\gamma \tilde{Y}_i^{(k)})}, \quad k = 1, 2, \dots, M_2. \tag{8}$$

(3) Resample from S_i according to the probabilities $\omega_{i1}(\gamma), \dots, \omega_{iM_2}(\gamma)$ to obtain $Y_i^{(1)}, \dots, Y_i^{(M)}$. To ensure the convergence of the aforementioned process, it is crucial to have $M_2 \to \infty$ and $M/M_2 \to 0$.

The SIR-based quantile regression estimation equation is given by

$$\psi^{eei}(Y_i, \mathbf{Z}_i, \delta_i; \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma) = \delta_i \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}) + (1 - \delta_i) \frac{1}{M} \sum_{j=1}^{M} \psi(\mathbf{Z}_i, Y_i^{(j)}(\boldsymbol{\beta}, \gamma); \boldsymbol{\theta}).$$

Due to the nonsmoothness of the aforementioned estimation equation, obtaining the sandwich estimator of the asymptotic covariance matrix directly is not feasible. Therefore, this paper proposes using a smooth function $\mathcal{G}(\mathbf{Z}_i^\top \boldsymbol{\theta} - Y_i)$ as a substitute for the indicator function $I(Y_i - \mathbf{Z}_i^\top \boldsymbol{\theta} < 0)$ in the quantile estimation equation, thus resulting in a smooth approximation of $\psi_\tau(Y_i - \mathbf{Z}_i^\top \boldsymbol{\theta})$:

$$\psi_h(Y_i, \mathbf{Z}_i; \boldsymbol{\theta}) = \mathbf{Z}^\top \mathcal{G}(\mathbf{Z}_i^\top \boldsymbol{\theta} - Y_i) - \tau,$$

where $\mathcal{G}_h(u) = \mathcal{G}(u/h)$, $\mathcal{G}(u) = \int_{-\infty}^{u} K(v) dv$, and $K(\cdot)$ is a kernel function defined in the range $[-1, 1]$.

For nonignorable missing data, we have the following representation for the smoothed SIR-based quantile regression estimation equation:

$$\psi_h^{eei}(Y_i, \mathbf{Z}_i, \delta_i; \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma) = \delta_i \psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}) + (1 - \delta_i) \frac{1}{M} \sum_{j=1}^{M} \psi_h(\mathbf{Z}_i, Y_i^{(j)}(\boldsymbol{\beta}, \gamma); \boldsymbol{\theta}).$$

The estimation equation based on imputation is susceptible to the misspecification of $f(Y|\mathbf{Z}, \delta = 1; \boldsymbol{\beta})$. Due to the relative robustness of the semiparametric response model, we consider the AIPW (augmented inverse probability weighting) estimation equation:

$$\psi_h^{aipw}(Y_i, \mathbf{Z}_i, \delta_i; \boldsymbol{\theta}, \boldsymbol{\beta}, \gamma) = \frac{\delta_i \mathbf{Z}_i \psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta})}{\pi(\mathbf{X}_i, Y_i; \hat{g}_\gamma; \gamma)}$$
$$+ \left(1 - \frac{\delta_i}{\pi(\mathbf{X}_i, Y_i; \hat{g}_\gamma; \gamma)} \right) \cdot \frac{1}{M} \sum_{j=1}^{M} \psi_h(\mathbf{Z}_i, Y_i^{(j)}(\boldsymbol{\beta}, \gamma); \boldsymbol{\theta}),$$

where it can be proven that the AIPW estimation equation is consistent in the case of the misspecification of the parameter model $f(Y|\mathbf{Z}, \delta = 1; \boldsymbol{\beta})$.

In practice, $(\boldsymbol{\beta}, \gamma)$ are often unknown and need to be estimated separately. The maximum likelihood estimation of $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}$, is the solution to the following score function:

$$\sum_{i=1}^{n} \delta_i \frac{\partial \ln f(y_i | \mathbf{x}_i, \delta_i = 1; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0. \tag{9}$$

Then, we consider the estimation of the tilting parameter γ. The semiparametric missing propensity model is analyzed by considering two estimation approaches for the tilting parameter γ: the profile two-step generalized method of the moments estimation and the kernel regression estimation for the nonparametric component $g(\cdot)$. To estimate the skewness parameter γ, we define the profile estimation equation as follows:

$$\xi(\mathbf{Z}_i, Y_i, \delta_i; g_\gamma, \gamma) = \left\{ \frac{\delta_i}{\pi(\mathbf{X}_i, Y_i; g_\gamma, \gamma)} - 1 \right\} h(\mathbf{V}_i) := \xi_i(g_\gamma, \gamma), \tag{10}$$

where $h(V)$ is an arbitrary specified function of the instrumental variable V, and $g_\gamma(\cdot)$ satisfies the following:

$$\exp(-g_\gamma(X_i)) = \frac{E(1 - \delta_i | X_i)}{E\{\delta_i \exp(\gamma Y_i) | X_i\}}.$$

Under the assumption of a correctly specified missingness mechanism, it holds that $E\{\xi_i(g_\gamma, \gamma)\} = \mathbf{0}$, and the vector $\xi_i(g_\gamma, \gamma)$ is overidentified with respect to γ. The profile two-stage generalized method of the moments estimation for γ is given by the following:

$$\hat{\gamma} = \arg\min_{\gamma \in \mathbb{R}} \bar{\xi}(\hat{g}_\gamma, \gamma)^\top W_n^{-1} \bar{\xi}(\hat{g}_\gamma, \gamma),$$

where $\bar{\xi}(\hat{g}_\gamma, \gamma) = \frac{1}{n} \sum_{i=1}^n \xi(Z_i, Y_i, \delta_i; \hat{g}_\gamma, \gamma)$, and $W_n = \frac{1}{n} \sum_{i=1}^n \xi(Z_i, Y_i, \delta_i; \hat{g}_\gamma, \gamma)^{\otimes 2}$. The estimator $\hat{g}_\gamma(\cdot)$ represents the kernel regression estimate of $g(\cdot)$ and satisfies the following equation:

$$\exp(-\hat{g}_\gamma(\hat{X}_i)) = \frac{\sum_{j=1}^n (1 - \delta_j) \mathcal{K}_h(\hat{X}_j - \hat{X}_i)}{\sum_{j=1}^n \delta_j \exp(\gamma Y_i) \mathcal{K}_h(\hat{X}_j - \hat{X}_i)},$$

where $\mathcal{K}_h(u_1, \ldots, u_d)$ represents the d-dimensional kernel function with a bandwidth h.

Define $\hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma})$ and $l = 1, 2$, which satisfy the following:

$$\hat{\psi}_{hi}^{(1)}(\theta, \hat{\beta}, \hat{\gamma}) = \delta_i \psi_h(Z_i, Y_i; \theta) + (1 - \delta_i) \hat{m}_{\psi_h}^0(Z_i; \theta, \hat{\beta}, \hat{\gamma}),$$

$$\hat{\psi}_{hi}^{(2)}(\theta, \hat{\beta}, \hat{\gamma}) = \frac{\delta_i}{\pi(X_i, Y_i; \hat{g}_{\hat{\gamma}}; \hat{\gamma})} \psi_h(Z_i, Y_i; \theta) + \left(1 - \frac{\delta_i}{\pi(X_i, Y_i; \hat{g}_{\hat{\gamma}}; \hat{\gamma})}\right) \hat{m}_{\psi_h}^0(Z_i; \theta, \hat{\beta}, \hat{\gamma}),$$

where $\hat{m}_{\psi_h}^0(Z_i; \theta, \hat{\beta}, \hat{\gamma}) = \frac{1}{M} \sum_{j=1}^M \psi_h(Z_i, Y_i^{*(j)}(\hat{\beta}, \hat{\gamma}); \theta)$.

Let p_i represent the probability mass of $\hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma})$, where $i = 1, 2, \cdots, n$. The empirical log-likelihood ratio function with respect to θ is defined as follows:

$$\hat{R}^{(l)}(\theta) = -2 \sup\left\{\sum_{i=1}^n \log(np_i) \middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma}) = 0\right\}.$$

Using the method of Lagrange multipliers, it can be shown that $\hat{R}^{(l)}(\theta)$ can be expressed as follows:

$$\hat{R}^{(l)}(\theta) = 2 \sum_{i=1}^n \log\{1 + \lambda^\top \hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma})\}, \tag{11}$$

where λ satisfies the following:

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma})}{1 + \lambda^\top \hat{\psi}_{hi}^{(l)}(\theta, \hat{\beta}, \hat{\gamma})} = 0.$$

The empirical likelihood estimators of the quantile regression coefficients based on the two proposed estimation equations in this paper, denoted as $\hat{\theta}^{(l)}$, $l = 1, 2$, are given by the following:

$$\hat{\theta}^{(l)} = \arg\min_{\theta} \hat{R}^{(l)}(\theta).$$

3. Theoretical Analysis

To elucidate the theoretical properties of the proposed estimators in this paper, we first define the matrix as follows:

$$A_1 = E\left\{ \pi(\boldsymbol{X}, Y)\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta}_0)^{\otimes 2} + (1 - \pi(\boldsymbol{X}, Y))m_\psi^0(\boldsymbol{Z}; \boldsymbol{\theta}_0)^{\otimes 2} \right\},$$

$$A_2 = E\left\{ \pi(\boldsymbol{X}, Y)^{-1}\{\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta}_0) - m_\psi^0(\boldsymbol{Z}; \boldsymbol{\theta}_0)\}^{\otimes 2} + m_\psi^0(\boldsymbol{Z}; \boldsymbol{\theta}_0)^{\otimes 2} \right\},$$

$$B_1 = \mathrm{Var}(\boldsymbol{H}_1\boldsymbol{\Psi}(\boldsymbol{\beta}_0) + \boldsymbol{H}_2\boldsymbol{\Phi}(\gamma_0)), B_2 = \mathrm{Var}(\boldsymbol{H}_3\boldsymbol{\Phi}(\gamma_0)),$$

$$\boldsymbol{T}_l = \boldsymbol{A}_l + \boldsymbol{B}_l, \quad l = 1, 2,$$

$$\boldsymbol{H}_1 = E\{(1 - \delta)\mathrm{Cov}_0(\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta}_0), s(\boldsymbol{Z}, Y; \boldsymbol{\beta}_0)|\boldsymbol{Z})\},$$

$$\boldsymbol{H}_2 = E\{(1 - \delta)\mathrm{Cov}_0(\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta}_0), Y|\boldsymbol{Z})\},$$

$$\boldsymbol{H}_3 = E\left\{ (1 - \delta)(Y - m_Y^0(\boldsymbol{X}))(\psi(\boldsymbol{Z}, Y; \boldsymbol{\theta}_0) - m_\psi^0(\boldsymbol{Z}; \boldsymbol{\theta}_0)) \right\}.$$

(C1)   (a) The density of $\boldsymbol{Z}$ is bounded and has continuous and bounded second-order derivatives; (b) the density of $\boldsymbol{Z}$ and the propensity $\pi(\boldsymbol{Z}, Y, \gamma_0)$ are bounded away from 0 ; and (c) $E\{\exp(2\gamma Y)\}$ is finite and $E\{\pi(\boldsymbol{Z}, Y, \gamma_0) \mid \boldsymbol{Z}\} \neq 1$ almost surely.

(C2)   Let $\mathcal{K}_h(\boldsymbol{u}) = \mathcal{K}(\boldsymbol{u}/h)/h^d$ and K denote a generic notation for a d-dimensional kernel function, and the value of d is determined by the context of use. $\mathcal{K}$ is a bounded, uniformly continuous, symmetric function of the m'th order satisfying the following conditions: $\int \mathcal{K}(\boldsymbol{s})d\boldsymbol{s} = 1$, $\boldsymbol{s} = (s_1, \ldots, s_d)$, $\int s_l^t \mathcal{K}(\boldsymbol{s})d\boldsymbol{s} = 0$, and $\int s_l^m \mathcal{K}(\boldsymbol{s})d\boldsymbol{s} \neq 0$ for any $l = 1 \ldots, d$ and $t = 1, \ldots, m - 1$.

(C3)   The bandwidth sequence h satisfies $nh^{2d} \to \infty$, $nh^d / \log n \to \infty$, and $nh^{2m} \to 0$ as $n \to \infty$; the order m satisfies $m \geq 2$ and $2m > d$.

(C4)   Let $\boldsymbol{W}(\gamma) = E\{\xi(g_\gamma, \gamma)^{\otimes 2}\}$ and $\boldsymbol{\Xi}(\gamma) = E\{(1 - \delta)(h(\boldsymbol{V}) - m_V^0(\boldsymbol{X}, \gamma)(Y - m_Y^0(\boldsymbol{X}, \gamma))\}$, $\boldsymbol{\Lambda}(\boldsymbol{Z}, Y, \delta, \gamma) = \left(\frac{\delta}{\pi(\boldsymbol{X}, Y)} - 1\right)(h(\boldsymbol{V}) - m_V^0(\boldsymbol{X}, \gamma))$, where $m_V^0(\boldsymbol{X}, \gamma) = E\{h(\boldsymbol{V})|\boldsymbol{X}, \delta = 0; \gamma\}$, and $m_Y^0(\boldsymbol{X}, \gamma) = E\{Y|\boldsymbol{X}, \delta = 0; \gamma\}$:

$$\boldsymbol{\Phi}(\boldsymbol{Z}, Y, \delta, \gamma) = -\left(\boldsymbol{\Xi}(\gamma)^\top \boldsymbol{W}(\gamma)^{-1}\boldsymbol{\Xi}(\gamma)\right)^{-1}\boldsymbol{\Xi}(\gamma)^\top \boldsymbol{W}(\gamma)^{-1}\boldsymbol{\Lambda}(\boldsymbol{Z}, Y, \delta, \gamma),$$

$E\{\boldsymbol{\Phi}_i(\gamma_0)\}^2 < \infty, \partial\boldsymbol{\Phi}_i(\gamma)/\partial\gamma$ exists at γ_0, $E\left\{\sup_\gamma \|\xi(g_\gamma, \gamma)\|\right\} < \infty$, γ_0 is the unique solution to $E\{\xi(g_\gamma, \gamma)\} = 0$, and $\boldsymbol{W}(\gamma_0)$ is positive definite.

(C5) $\{(\boldsymbol{Z}_i, Y_i, \delta_i) : i = 1, \ldots, n\}$ are independent and identically distributed random vectors. The support of $\boldsymbol{\theta}$ denoted by $\mathcal{B}$ is a compact set in R^q, and $\boldsymbol{\theta}_0 \in \mathcal{B}$ is the unique solution to $E\{\psi(\boldsymbol{Z}_i, Y_i, \boldsymbol{\theta})\} = 0$. Furthermore, $\|\partial\psi(\boldsymbol{Z}_i, Y_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\|$, $\left\|\partial^2\psi(\boldsymbol{Z}_i, Y_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\right\|$ and $\|\psi(\boldsymbol{Z}_i, Y_i, \boldsymbol{\theta})\|^3$ are bounded by an integrable function $H(x, y)$ within a neighborhood of $\boldsymbol{\theta}_0$.

(C6)   For all ϵ in a neighborhood of zero and for almost every $\boldsymbol{Z}$, $F(\epsilon \mid \boldsymbol{Z})$, $f(\epsilon \mid \boldsymbol{Z})$ and $f(\epsilon \mid \boldsymbol{Z}, \delta = 0)$ to exist, they are bounded away from zero and are r times continuously differentiable with $r \geq 2$. There exists a function $C(\boldsymbol{Z})$ such that $\left|f^{(s)}(\epsilon \mid \boldsymbol{Z})\right|$, and $\left|f^{(s)}(\epsilon \mid \boldsymbol{Z}, \delta = 0)\right| \leq C(\boldsymbol{Z})$ for $s = 0, 2, \ldots, r$, for almost all $\boldsymbol{Z}$ and ϵ in a neighborhood of zero, and $E\left[C(X)\|X\|^2\right] < \infty$.

(C7)   The kernel function $K(\cdot)$ is a probability density function such that (a) it is bounded and has a compact support; (b) $K(\cdot)$ is an rth order kernel, i.e., $K(\cdot)$ satisfies $\int u^j K(u)du = 1$ if $j = 0$; 0 if $1 \leq j \leq r - 1$, and C_K if $j = r$ for some constant $C_k \neq 0$; and (c) we let $\tilde{G}(u) = \left(G(u), G^2(u), \ldots, G^{L+1}(u)\right)$ for some $L \geq 1$, where $G(u) = \int_{\tilde{v} < u} K(v)dv$. For any $\boldsymbol{\iota} \in R^{L+1}$ satisfying $\|\boldsymbol{\iota}\| = 1$, there is a partition of $[-1, 1], -1 = a_0 < a_1 < \cdots a_{L+1}$ such that $\boldsymbol{\iota}^\top \tilde{G}(u)$ is either strictly positive or strictly negative within (a_{l-1}, a_l) for $l = 1, \ldots, L + 1$.

(C8)   The positive bandwidth parameter h satisfies $nh^{2r} \to 0$ and $nh/\log(n) \to \infty$ as $n \to \infty$.

(C9) $\boldsymbol{Z}$ has a bounded support, $E\|\boldsymbol{Z}\|^4 < \infty$ and the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{T}_l$; additionally, $l = 1, 2$ are nonsingular.

(C10) Under complete observation of $(\mathbf{Z}_i, Y_i)$ for $i = 1, \dots, n$, the unique solution $\hat{\beta}$ to the score equation in (3) satisfies

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} N(0, \Sigma),$$

for some Σ for sufficiently large n.

To ensure the requirements of Lemma 8.11 by Newey and McFadden [22] and Theorem 6.18 by Van [23], the conditions (C1)–(C4), which are commonly found in the literature on missing data and nonparametric method [24,25], are primarily employed. These conditions encompass the following: (1) random equivalence and continuity conditions; (2) linearity conditions on the objective function with respect to nonparametric components and convergence rate conditions for nonparametric estimators; and (3) the differential continuity condition of the estimating equations with respect to the parameter of interest. Conditions (C5)–(C9) ensure the consistency and asymptotic normality of the empirical likelihood estimator for quantile regression smoothing [10]. To simplify the discussion on the asymptotic properties of maximum likelihood estimation in the working model, we introduce condition (C10).

Under the fulfillment of the assumed conditions, we define the following:

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{n}} \Phi_i(\gamma_0) + o_p(1),$$

$$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \Psi_i(\beta_0) + o_p(1).$$

In addition, we have the following lemma, whose proof is given in the Appendix A:

Lemma 1. *Under conditions (C5)–(C9), we have*

$$E\{\psi_h(\mathbf{Z}_i, Y_i; \theta_0)\} = O(h^r),$$
$$E\{m_{\psi_h}^0(\mathbf{Z}_i; \theta_0)\} = E\{m_{\psi}^0(\mathbf{Z}_i; \theta_0)\} + O(h^r).$$

Lemma 2. *Under the assumption conditions (C1)–(C10), with the notation from Section 3, the following results hold as $n \to \infty$:*

$$(1) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{hi}^{(l)}(\theta_0, \hat{\beta}, \hat{\gamma}) \xrightarrow{d} N(\mathbf{0}, \mathbf{T}_l); \qquad (2) \frac{1}{n} \sum_{i=1}^{n} \psi_{hi}^{(l)}(\theta_0, \hat{\beta}, \hat{\gamma}) \xrightarrow{p} \mathbf{V}_l;$$

$$(3) \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \psi_{hi}^{(l)}(\theta_0, \hat{\beta}, \hat{\gamma})}{\partial \theta^\top} \xrightarrow{p} \Gamma; \qquad (4) \max_{i} \|\psi_{hi}^{(l)}(\theta_0, \hat{\beta}, \hat{\gamma})\| = o_p(n^{1/2}).$$

Theorem 1. *Under conditions (C1)–(C10), if the parameter working model is correctly specified, as $n \to \infty$ for $l = 1, 2$, we have*

$$\sqrt{n}\left(\hat{\theta}^{(l)} - \theta_0\right) \xrightarrow{d} N\left(0, \Gamma^{-1}\mathbf{T}_l\Gamma^{-\top}\right),$$

where $\Gamma = E\left\{f(0 \mid \mathbf{Z})\mathbf{Z}\mathbf{Z}^\top\right\}$.

If there is no missing data, we have $P(\delta|\mathbf{Z}, Y) = 0$, which implies that $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ are all zero. Additionally, we have

$$\mathbf{A}_1 = \mathbf{A}_2 = E\left\{\psi(\mathbf{Z}, Y, \theta_0)^{\otimes 2}\right\} = \tau(1 - \tau)E\left\{\mathbf{Z}_i\mathbf{Z}_i^\top\right\}.$$

The above results are consistent with the asymptotic normality conclusion of classical quantile regression.

The different forms of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ indicate that if the parameter working model $f(Y|\mathbf{Z}, \delta = 1; \boldsymbol{\beta})$ is misspecified, $\hat{\boldsymbol{\theta}}^{(1)}$ is no longer consistent, while $\hat{\boldsymbol{\theta}}^{(2)}$ remains a consistent estimator of $\boldsymbol{\theta}_0$. The following procedure demonstrates the double robustness of the AIPW estimation equation. For misspecified $f(Y|\mathbf{Z}, \delta = 1)$ values, there exists $C(\mathbf{Z}; \boldsymbol{\theta}, \hat{\gamma})$ such that

$$\hat{m}_{\psi}^0(\mathbf{Z}_i; \boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma}) = C(\mathbf{Z}_i; \boldsymbol{\theta}, \hat{\gamma}) + o_p(1).$$

This illustrates the double robustness property of the AIPW estimation equation.

It can be shown that

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta})}{\pi(\mathbf{X}_i, Y_i; \hat{g}_{\hat{\gamma}}, \hat{\gamma})} \\
&\quad + \left\{ 1 - \frac{\delta_i}{\pi(\mathbf{X}_i, Y_i; \hat{g}_{\hat{\gamma}}, \hat{\gamma})} \right\} C(\mathbf{Z}; \boldsymbol{\theta}, \hat{\gamma}) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta})}{\pi(\mathbf{X}_i, Y_i; \hat{g}_{\hat{\gamma}}, \hat{\gamma})} + o_p(1) \\
&=: \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\psi}_{hi}^{ipw}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma}) + o_p(1).
\end{aligned}
$$

If $\pi(\mathbf{X}_i, Y_i)$ is correctly specified, the IPW estimation equation is consistent, which implies that the AIPW quantile regression estimation equation remains consistent in this case.

Theorem 2. *Under the conditions of Theorem 1, if the parameter working model is correctly specified, for $l = 1, 2$ and as $n \to \infty$,*

$$\hat{R}^{(l)}(\boldsymbol{\theta}_0) \xrightarrow{d} r_1^{(l)} \chi_{1 \cdot 1}^2 + r_2^{(l)} \chi_{1 \cdot 2}^2 + \cdots + r_q^{(l)} \chi_{1 \cdot q}^2,$$

where $r_i^{(l)}$ are the eigenvalues of $\mathbf{A}_l^{-1} \boldsymbol{\Lambda}_l$. $\chi_{1 \cdot 1}^2, \chi_{1 \cdot 2}^2, \ldots, \chi_{1 \cdot q}^2$ represent q independent standard χ^2 distributed random variables.

First, if there is no data missingness, we have $\mathbf{T}_l = \mathbf{A}_l = \tau(1 - \tau)E\left\{ \mathbf{Z}\mathbf{Z}^\top \right\}$, which leads to $\hat{R}^{(l)}(\boldsymbol{\theta}) \xrightarrow{d} \chi_q^2$, and the Wilks' Theorem holds. Furthermore, it is worth noting that if $\boldsymbol{\beta}$ and γ are known, we still have $\mathbf{T}_l = \mathbf{A}_l$, and in this case, the Wilks' Theorem still holds. The above conclusion is consistent with Zhao [26].

4. Simulation Study

To investigate the finite-sample properties of the proposed method, this study conducted numerical simulations under both correctly specified and misspecified working model scenarios.

4.1. Simulation 1: Correctly Specified Working Model

In the numerical simulation, we generated a random vector (x, y, δ), where x is the independent variable, y is the response variable of interest, and δ is the indicator variable for the observation of y. When $\delta = 1$, y has observed values; otherwise, the observation of y is missing. Let $x \sim N(0, 0.5)$, and generate the observed response variable y according to the following equation:

$$y = \mu(x) + e,$$

where $\mu(x) = -1 + x$. We considered two different distributions for the random error term e: (a) $N(0, 0.9)$ and (b) $N(0, 0.49(1 + x^2))$. For the working model $f(y|x, \delta = 1; \boldsymbol{\beta})$, the former follows a homoscedastic structure $N(\mu(x), \sigma_1^2)$, while the latter follows a heteroscedastic structure $N(\mu(x), \sigma_2^2(x))$.

The indicator variable δ follows a Bernoulli distribution with parameter p, i.e., $\delta \sim$ Bernoulli$(1, p)$. The conditional probability of δ given (x, y) is defined as follows:

$$p(\boldsymbol{\phi}) = P(\delta = 1 | x, y) = \{1 + \exp(-\phi_0 - \phi_1 y)\}^{-1},$$

where $(\phi_0, \phi_1) = (0.8, -0.2)$. In this case, the missingness mechanism for the response variable y is nonrandom, with x serving as a missingness instrument. The average observed rate in the sample was approximately 73%.

We establish a quantile regression model of the response variable y on x as follows:

$$Q_\tau(y|x) = \theta_0 + \theta_1 x,$$

where τ represents the quantile of interest, specifically $\tau = (0.25, 0.5, 0.75)$.

Consider the following five quantile regression estimation equations:

(1) Full Estimation Equation: $\psi_h^{\text{full}}(x_i, y_i; \boldsymbol{\theta}) = \psi_h(x_i, y_i; \boldsymbol{\theta})$;
(2) Complete Case (CC) Estimation Equation: $\psi_h^{\text{cc}}(x_i, y_i, \delta_i; \boldsymbol{\theta}) = \delta_i \psi_h(x_i, y_i; \boldsymbol{\theta})$;
(3) IPW Estimation Equation: $\psi_h^{\text{ipw}}(x_i, y_i, \delta_i; \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})$;
(4) EEI Estimation Equation: $\psi_h^{\text{eei}}(x_i, y_i, \delta_i; \boldsymbol{\theta}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$;
(5) AIPW Estimation Equation: $\psi_h^{\text{aipw}}(x_i, y_i, \delta_i; \boldsymbol{\theta}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$.

To generate a sample of size $n = 500$ that meets the requirements of the simulation, we can use the law of total probability and express $f(y|x)$ as follows:

$$f(y|x) = P(\delta = 1|x)f(y|x, \delta = 1) + P(\delta = 0|x)f(y|x, \delta = 0),$$

where

$$f(y|x, \delta = 0) = f(y|x, \delta = 1) \times \frac{O(x, y)}{E(O(x, Y)|x, \delta = 1)}.$$

Under the specified nonrandom missingness mechanism, we have $O(x, y) = \exp(-\phi_0 - \phi_1 y)$ for the homoscedastic case of $f(y|x, \delta = 1) = N(\mu, \sigma_1^2)$. In this case, we can express the ratio of the conditional probabilities as follows:

$$\frac{P(\delta = 0|x)}{P(\delta = 1|x)} = E(O(x, y)|x, \delta = 1) = \exp(\phi_0 - \phi_0'),$$

where $\phi_0' = -\frac{1}{2\sigma_1^2}(2\mu\sigma_1^2\phi_1 + \sigma_1^4\phi_1^2)$. Thus, we have $f(y|x, \delta = 0) = N(\mu + \sigma_1^2\phi_1, \sigma_1^2)$ and $P(\delta = 1|x) = 1 + \exp(\phi_0 - \phi_0')^{-1}$.

Since x is completely observed, we can draw a sample of size $n = 500$ from the mixed distribution of $f(y|x)$. A similar approach can be applied under the heteroscedastic assumption.

It should be noted that the response variable y originates from a distribution with a complex, mixed form. As a result, discussing the true values of the parameters (θ_0, θ_1) poses a formidable challenge. This complexity renders it difficult to assess the performance of the estimation methods using conventional measures such as bias or the root mean squared error (RMSE). Consequently, we introduce the following approxmate relative evaluation metrics:

$$ARE(Method*, Full) = \frac{ARMSE(Method*)}{ARMSE(Full)},$$

where $ARMSE(Method*) = \sqrt{SD(Method*)^2 + (Mean(Method*) - Mean(Full))^2}$.

Tables 1 and 2 summarize the mean and variance of the five coefficient estimates at different quantiles based on 1000 Monte Carlo simulations under the homoscedastic case (a) and heteroscedastic case (b) of $f(y|x, \delta = 1)$. From the estimation results, it can be observed that the coefficient estimates based on complete observations have larger bias compared to the other estimation methods. When the working model $f(y|x, \delta = 1; \boldsymbol{\beta})$ was correctly

specified, the proposed imputation estimates yielded smaller variances compared to the IPW estimates. In this case, the performance of the AIPW estimates was similar to the IPW estimates. Comparing the results at different quantiles, it can be seen that the variances of the five estimation methods at the 0.5 quantile were smaller than those at the 0.25 and 0.75 quantiles, which is due to the larger sample size at the central quantile compared to the tails. Under the homoscedastic assumption, the variances of the estimates at the 0.25 and 0.75 quantiles were similar. Under the existing missing mechanism, as the value of the response variable y increased, the missing propensity also increased, thereby indicating higher missing rates at the upper quantiles. Consequently, the estimation variances of the IPW and AIPW estimates were higher at the high quantile of $\tau = 0.75$ compared to the low quantile of $\tau = 0.25$. However, proper imputation could greatly improve the estimation efficiency at the high quantile of $\tau = 0.75$. This improvement was more pronounced under the heteroscedastic model. These results demonstrate that the imputation estimates are nearly unbiased when the working model is correctly specified and have higher estimation efficiency compared to the IPW and AIPW estimates.

Table 1. Monte Carlo mean, standard deviation (SD), and approximate relative performance (ARE) of the five methods for error term (a).

		FULL		CC		EEI		IPW		AIPW	
		θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1
$\tau = 0.25$	Mean	-1.606	1.009	-1.652	1.002	-1.598	1.008	-1.605	1.008	-1.608	1.008
	SD	0.054	0.105	0.063	0.123	0.049	0.095	0.062	0.121	0.061	0.122
	ARE	1.000	1.000	1.445	1.173	0.919	0.905	1.148	1.152	1.130	1.162
$\tau = 0.5$	Mean	-0.950	1.010	-0.999	1.002	-0.953	1.008	-0.950	1.009	-0.950	1.008
	SD	0.049	0.098	0.057	0.116	0.044	0.089	0.057	0.117	0.057	0.118
	ARE	1.000	1.000	1.534	1.186	0.900	0.908	1.163	1.194	1.163	1.204
$\tau = 0.75$	Mean	-0.298	1.009	-0.349	1.001	-0.310	1.008	-0.297	1.008	-0.293	1.008
	SD	0.052	0.107	0.061	0.126	0.046	0.097	0.063	0.130	0.062	0.131
	ARE	1.000	1.000	1.529	1.180	0.914	0.907	1.212	1.215	1.196	1.224

Table 2. Monte Carlo mean, standard deviation (SD), and approximate relative performance (ARE) of the five methods for error term (b).

		FULL		CC		EEI		IPW		AIPW	
		θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1
$\tau = 0.25$	Mean	-1.503	1.004	-1.534	1.004	-1.492	1.005	-1.502	1.006	-1.507	1.005
	SD	0.045	0.099	0.052	0.112	0.041	0.088	0.051	0.109	0.051	0.110
	ARE	1.000	1.000	1.345	1.131	0.943	0.889	1.134	1.101	1.137	1.111
$\tau = 0.5$	Mean	-0.968	1.004	-1.000	0.0997	-0.970	1.003	-0.968	1.004	-0.968	1.003
	SD	0.040	0.092	0.046	0.104	0.036	0.081	0.046	0.105	0.047	0.106
	ARE	1.000	1.000	1.401	1.133	0.901	0.881	1.150	1.141	1.175	1.152
$\tau = 0.75$	Mean	-0.433	1.006	-0.467	0.996	-0.447	1.007	-0.432	1.007	-0.428	1.006
	SD	0.044	0.101	0.051	0.119	0.039	0.090	0.053	0.124	0.052	0.123
	ARE	1.000	1.000	1.392	1.182	0.942	0.891	1.205	1.228	1.187	1.218

4.2. Simulation 2: Misspecification of the Working Model

In practical situations, the true data generation mechanism is unknown, and it is challenging to accurately specify the working model $f(y|x, \delta = 1; \beta)$ for the observed data. In this study, we investigated the finite sample properties of the proposed imputation estimator and calibration estimator under the misspecification of the working model. The simulation model includes two covariates: $X_1 \sim N(0, 1)$ and $X_2 \sim \text{Exp}(0.2)$. Given the covariates, the response variable Y is generated as follows:

$$Y = 1 + X_1 + X_2 + 0.25(2 + X_2)\varepsilon,$$

where $\varepsilon \sim N(0,1)$, and X_1, X_2, and ε are mutually independent.

In this simulation setup, the error term distribution of Y is heteroscedastic. The missing data mechanism for Y is nonrandom and follows

$$P(R = 1|\boldsymbol{X}, Y) = \frac{1}{1 + \exp(-0.1 + 0.5X_1^2 + 0.15Y)}.$$

The average observed rate in the model was approximately 73%, and X_2 served as an instrumental variable. We generated a random sample of size $n = 500$ denoted as $(\boldsymbol{X}_i, Y_i, \delta_i) : i = 1, \ldots, n$. For the aforementioned simulation model, we consider the following quantile regression model:

$$Q_\tau(Y|\boldsymbol{X}) = (1, \boldsymbol{X}^\top)\boldsymbol{\theta}_0(\tau),$$

where $\boldsymbol{\theta}_0(\tau) = (1 + 0.5Q_\tau(\varepsilon), 1, 1 + 0.25Q_\tau(\varepsilon))^\top$.

Under the aforementioned data generating mechanism, obtaining an explicit expression for $f(Y|\boldsymbol{X}, \delta = 1; \boldsymbol{\beta})$ is challenging and requires specifying the working model based on the observed data. In this simulation model, we consider two possible working models: (1) $N(\hat{\mu}(\boldsymbol{X}), \hat{\sigma}^2)$ and (2) $N(\hat{\mu}(\boldsymbol{X}), (0.5 + 0.25X_2)^2)$. Figures 1 and 2 illustrate that the residual distribution of the working model (1) exhibited peakedness, thus violating the normality assumption and indicating model misspecification. In contrast, working model (2) took into account the correct specification of the variance.
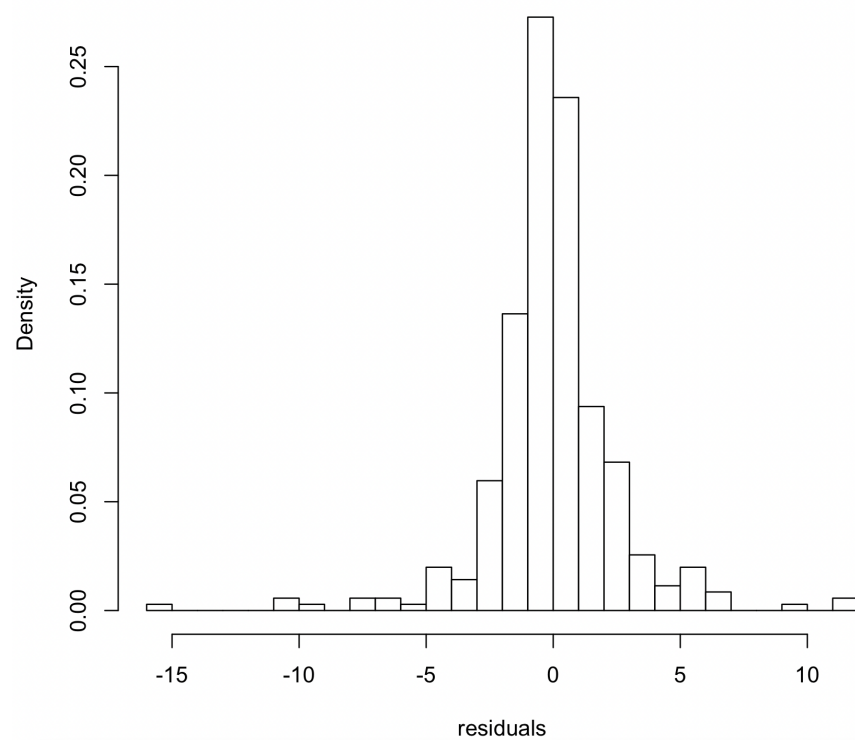

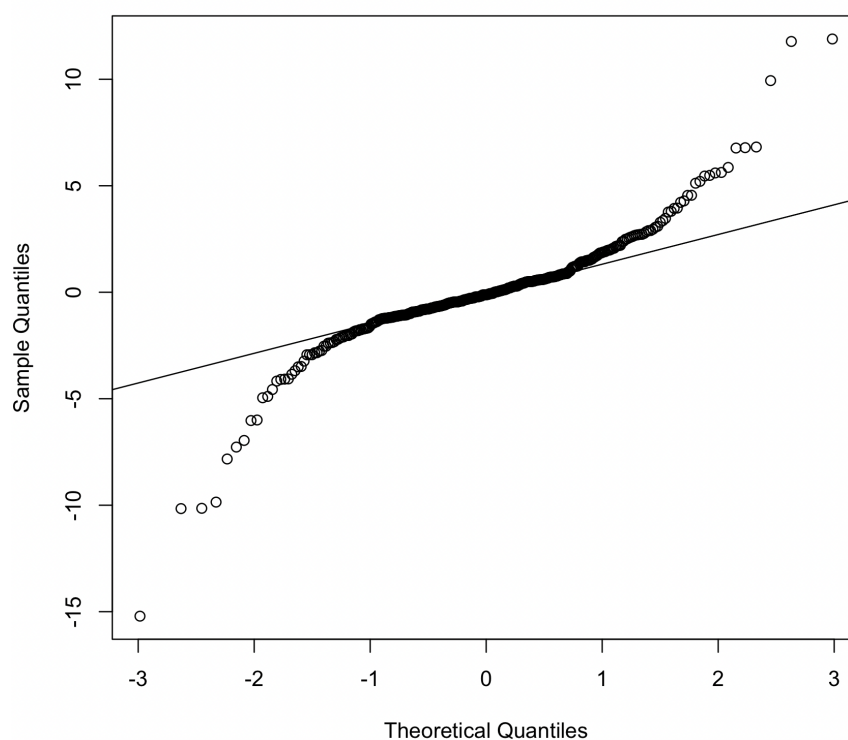**Figure 1.** Histogram of the residual distribution for the parameter working model $N(\hat{\mu}(\boldsymbol{X}), \hat{\sigma}^2)$.

Figure 2. QQ plot of the residual distribution for the parameter working model $N(\hat{\mu}(\boldsymbol{X}), \hat{\sigma}^2)$.

The estimation results of the five types of quantile regression estimates obtained from 1000 random simulations at different quantiles are summarized in Tables 3–5. These tables include two types of imputation estimates based on the parameter working models (1) and (2), as well as the corresponding AIPW estimates based on the parameter working models (1) and (2), and the combined estimation equations. The results show that the imputation estimates based on the erroneously specified working model (1) exhibited significant estimation bias. On the other hand, although the working model (2) was also misspecified, it took into account the heteroscedasticity in the conditional distribution of the response variable, thus resulting in smaller estimation bias compared to model (1) and better estimation performance. These findings highlight the sensitivity of imputation methods to misspecified working models. Across the three quantiles, the IPW estimates performed well, thus indicating the robustness of the semiparametric response assumption. Even in the presence of misspecified parameter working models, both of the AIPW estimates had similar median absolute deviations to IPW, which were significantly smaller than the misspecified imputation estimates, thereby demonstrating the robustness of the AIPW estimation. Comparing the two AIPW estimates, it is observed that the estimate based on the correctly specified parameter working model had smaller estimation bias and higher estimation efficiency.

Table 3. The bias (Bias), standard deviation (SD), and median absolute deviation (MAD) of the five types of quantile regression coefficient estimates at $\tau = 0.25$.

Method	θ_0			θ_1			θ_2		
	Bias	SD	MAD	Bias	SD	MAD	Bias	SD	MAD
full	−0.021	0.100	0.067	0.002	0.069	0.046	0.004	0.038	0.026
cc	0.001	0.122	0.086	−0.003	0.079	0.052	0.020	0.041	0.032
ipw	−0.039	0.149	0.081	−0.001	0.103	0.053	0.006	0.065	0.029
eei.0	−0.362	0.158	0.358	0.019	0.098	0.067	0.047	0.044	0.049
eei.1	−0.003	0.137	0.093	0.002	0.082	0.054	0.003	0.043	0.029
aipw.0	−0.026	0.120	0.086	0.002	0.081	0.054	0.005	0.041	0.028
aipw.1	−0.024	0.115	0.082	0.002	0.078	0.052	0.005	0.040	0.027

Table 4. The bias (Bias), standard deviation (SD), and median absolute deviation (MAD) of the five types of quantile regression coefficient estimates at $\tau = 0.5$.

Method	θ_0			θ_1			θ_2		
	Bias	SD	MAD	Bias	SD	MAD	Bias	SD	MAD
full	0.001	0.087	0.056	0.001	0.064	0.044	0.001	0.034	0.022
cc	0.157	1.089	0.076	−0.039	0.398	0.048	−0.003	0.151	0.027
ipw	0.029	0.103	0.067	−0.017	0.395	0.047	−0.001	0.065	0.024
eei.0	−0.089	0.118	0.104	0.006	0.076	0.048	0.013	0.039	0.026
eei.1	0.007	0.127	0.085	0.001	0.077	0.049	0.001	0.039	0.026
aipw.0	−0.008	0.159	0.066	0.001	0.110	0.047	0.002	0.037	0.025
aipw.1	−0.002	0.124	0.069	0.003	0.079	0.046	0.002	0.036	0.024

Table 5. The bias (Bias), standard deviation (SD), and median absolute deviation (MAD) of the five types of quantile regression coefficient estimates at $\tau = 0.75$.

Method	θ_0			θ_1			θ_2		
	Bias	SD	MAD	Bias	SD	MAD	Bias	SD	MAD
full	0.019	0.095	0.063	0.001	0.072	0.048	−0.001	0.037	0.025
cc	0.048	0.117	0.082	−0.003	0.080	0.054	0.009	0.041	0.029
ipw	0.011	0.246	0.070	0.002	0.079	0.051	0.001	0.049	0.026
eei.0	0.092	0.127	0.106	−0.003	0.084	0.055	−0.009	0.041	0.029
eei.1	0.019	0.131	0.084	−0.001	0.081	0.054	−0.002	0.0419	0.029
aipw.0	−0.047	0.510	0.073	0.020	0.361	0.052	0.001	0.072	0.026
aipw.1	−0.017	0.424	0.072	0.006	0.248	0.052	−0.001	0.064	0.025

5. Real Data Application

We applied our proposed method to the data of 2139 HIV-infected patients enrolled in the ACTG175 study [27]. The ACTG175 study evaluated the efficacy of monotherapy or combination therapy in HIV-infected patients with CD4 cell counts between 200 and 500 cells/mm³. Following the studies by Davidian et al. [28] and Zhang et al. [29], we categorized all the treatment regimens into two groups. The first group consisted of the standard zidovudine (ZDV) monotherapy arm, while the second group included three newer treatment arms: ZDV and dual nucleoside analogue (ddl), ZDV and zalcitabine (ddC), and ddl monotherapy. The first group comprised 532 subjects, while the second group comprised 1697 subjects. We investigated the effect of the treatment arm (trt, 0 = ZDV monotherapy only) on the τ quantile of the CD4 cell count ($CD4_{96}$) measured at baseline and adjusted for the baseline CD4 cell count ($CD4_0$) and other baseline covariates, including age, weight, race (0 = Caucasian), gender (0 = female), history of reverse transcriptase inhibitor use (0 = no), and whether the subject discontinued treatment before 96 weeks (offtrt, 0 = no).

Consider fitting a linear quantile regression model as follows:

$$Q_\tau(\text{CD4}_{96}|X) = \beta_1(\tau) + \beta_2(\tau)\text{trt} + \beta_3(\tau)\text{CD4}_0 + \beta_4(\tau)\text{age} + \beta_5(\tau)\text{weight}$$
$$+ \beta_6(\tau)\text{race} + \beta_7(\tau)\text{gender} + \beta_8(\tau)\text{history} + \beta_9(\tau)\text{offtrt}.$$

The dataset used in this study is sourced from the R package "speff2trial". The study population consists of 1522 Caucasian individuals and 617 non-Caucasian individuals, with 1171 males and 368 females. The average age of the participants is 35 years, with a standard deviation of 8.7 years. Among the participants, 1253 individuals had a history of antiretroviral therapy, and 776 individuals discontinued treatment before the 96th week.

Due to attrition during the study period, approximately 37% of the participants have missing values for the variable CD496. Although complete measurements of other variables related to CD496, such as baseline CD4 and CD8 cell counts CD40 and CD80, as well as CD4 and CD8 cell counts at 20 ± 5 weeks CD420 and CD820, were obtained at baseline and follow-up visits, these variables may not fully explain the propensity for participants to drop out. In other words, we cannot assume that the missingness of CD4_{96} is random. Therefore, in our analysis, we consider a more comprehensive semiparametric nonrandom missingness mechanism:

$$P(R = 1|X, S, Y) = \pi(S, Y) = \frac{1}{1 + \exp(g(S) + \gamma Y)},$$

where S represents the set of variables associated with attrition, and $g(S)$ is a function capturing the relationship between these variables and the missingness indicator R.

Figure 3 displays the histograms of the observed CD4_{96} and its logarithm. From the figure, it can be observed that the conditional distribution $f(y|X, R = 1)$ of observed CD4_{96} is right-skewed. However, the logarithmic transformation did not result in improved symmetry, thus indicating that the normality assumption did not hold. In our analysis, we can assume that CD4_{96} follows a truncated normal distribution with left truncation at 0, where its mean is primarily determined by the influence of eight covariates and three auxiliary variables.

The parameters β in the working model $f(y|X, R = 1; \beta)$ are estimated using the truncation regression model in R package "truncreg". The parameter γ in $\pi(S, Y)$ is estimated using the method of the profile generalized method of moments (GMMs).

Figures 4 and 5 illustrate the normality properties of the residuals from the truncated regression working model. Visually, the distribution of residuals appears to be symmetric. The calculated sample skewness is 0.05, thus indicating a slight deviation from perfect symmetry. The Q-Q plot reveals that the distribution of residuals has a kurtosis less than 3. Further computation reveals a kurtosis of 2.11, thus indicating that the residual distribution is flatter than a standard normal distribution.

Table 6 presents the estimates of the quantile regression coefficients and corresponding 95% confidence intervals at the $\tau = 0.25, 0.5, 0.75$ quantile levels. The four estimation methods considered include complete case (CC) estimation, inverse probability weighting (IPW) estimation, multiple imputation (MI) estimation, and augmented inverse probability weighting (AIPW) estimation. The MI estimation is based on averaging over $L = 20$ randomly generated imputations. Confidence intervals for the coefficient estimates were obtained using the bootstrap method with $B = 200$ resampling iterations.
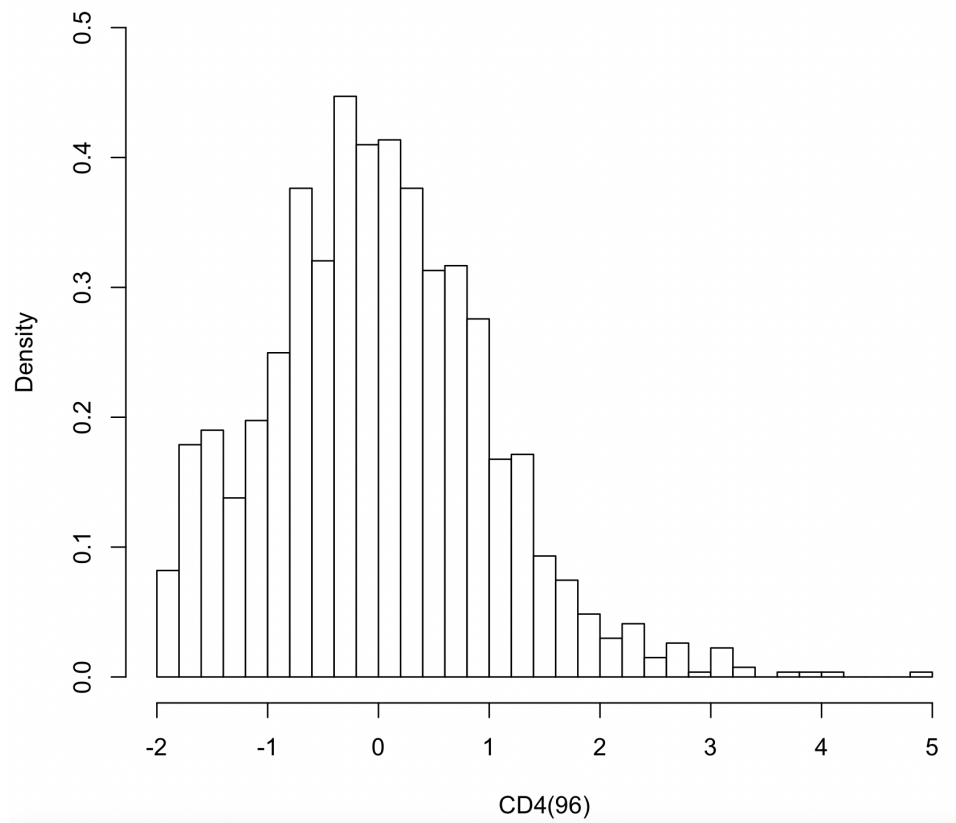
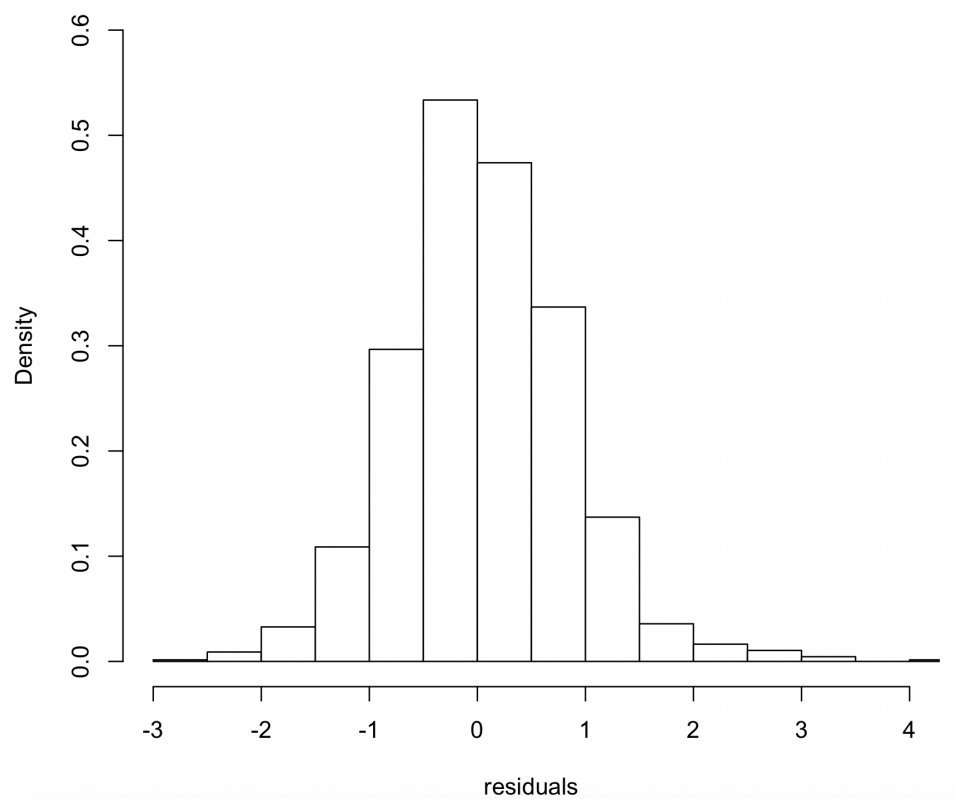**Figure 3.** Histogram of complete observed data CD4$_{96}$ in ACTG175.

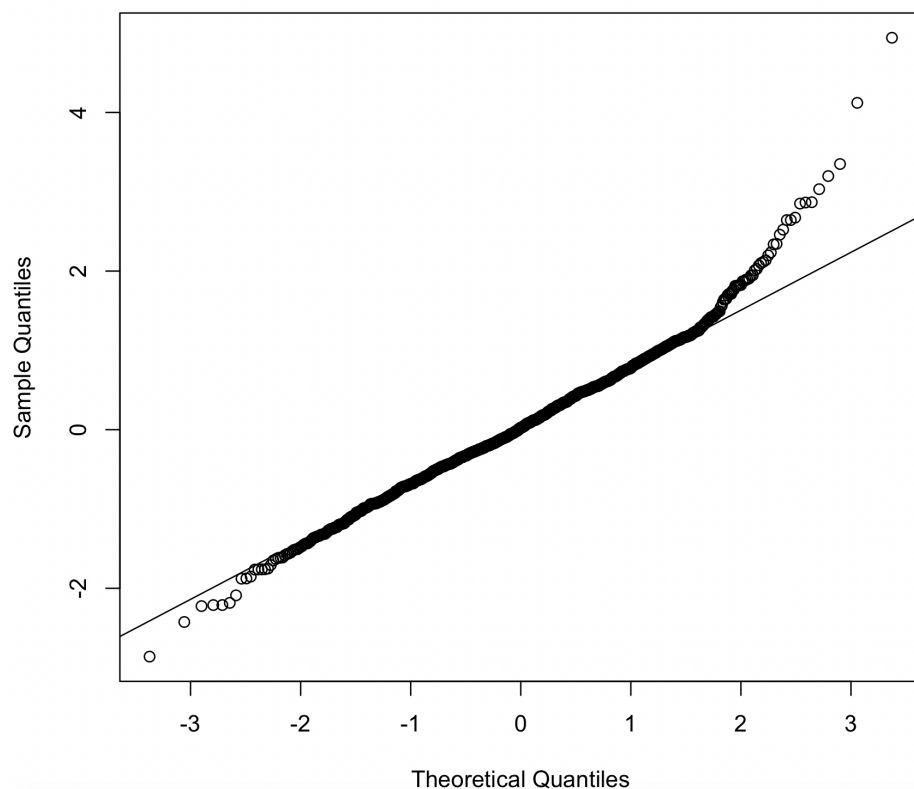**Figure 4.** Histogram of residuals from the parameterized working model.

Figure 5. Q-Q plot of residuals from the parameterized working model.

From Table 6, it can be observed that for the three given quantile levels and four estimation methods, patients receiving the three new combined treatment methods had significantly higher CD4 cell counts at 96 ± 5 weeks compared to the traditional treatment method. In other words, the new treatment methods had significantly slowed down the progression of AIDS compared to the traditional method. Comparing the four estimation methods, it is evident that the complete case estimation overestimated the performance of the treatment group. The results of the IPW estimation and AIPW estimation were similar and higher than the imputation estimation. When comparing the treatment effects at different quantile levels, both the IPW estimation and imputation estimation reflected a decreasing trend in treatment effect from the 0.25th to the 0.75th quantile. Although the AIPW estimation and complete case estimation did not show a similar trend, the coefficient estimates of the AIPW estimation also indicate a more significant improvement in treatment effect for patients at lower quantiles.

Upon examining the effects of the other covariates, it is found that for all four estimation methods, the baseline CD4 level $CD4_0$ had a positive impact on the CD4 cell count at 96 ± 6 weeks, while patients with a history of antiretroviral therapy or early treatment discontinuation exhibited poorer CD4 cell levels at 96 ± 5 weeks. In comparison to the covariates directly related to the disease progression mentioned above, the effects of age, weight, race, gender, and other covariates on the CD4 cell count at 96 ± 6 weeks were minimal. The impact directions and significance obtained from different methods were also not consistent. Therefore, although these variables needed to be considered in the modeling process, conclusions regarding their effects should be drawn with caution.

Table 6. Analysis results of the ACTG175 dataset.

Covariate	AIPW Estimator		IPW Estimator		EEI Estimator		CC Estimator	
	Est	CI	Est	CI	Est	CI	Est	CI
$\tau = 0.25$								
intercept	−0.527	(−0.576, −0.478)	−0.493	(−0.656, −0.362)	−0.528	(−0.708, −0.416)	−0.509	(−0.671, −0.386)
age	−0.001	(−0.019, 0.021)	0.001	(−0.058, 0.056)	0.091	(0.029, 0.141)	−0.006	(−0.061, 0.059)
wtkg	0.007	(−0.012, 0.026)	0.029	(−0.025, 0.083)	−0.147	(−0.198, −0.085)	0.031	(−0.025, 0.091)
race	−0.057	(−0.091, −0.022)	−0.089	(−0.218, 0.026)	−0.056	(−0.152, 0.032)	−0.092	(−0.204, 0.027)
gender	−0.002	(−0.042, 0.037)	−0.061	(−0.148, 0.075)	0.024	(−0.079, 0.176)	−0.057	(−0.131, 0.076)
history	−0.231	(−0.266, −0.197)	−0.234	(−0.349, −0.142)	−0.216	(−0.293, −0.123)	−0.227	(−0.331, −0.134)
offtrt	−0.549	(−0.584, −0.513)	−0.528	(−0.716, −0.416)	−0.399	(−0.449, −0.268)	−0.553	(−0.717, −0.434)
$CD4_0$	0.474	(0.459, 0.489)	0.496	(0.446, 0.549)	0.472	(0.416, 0.506)	0.493	(0.442, 0.543)
trt	0.367	(0.332, 0.401)	0.369	(0.254, 0.475)	0.239	(0.168, 0.344)	0.377	(0.261, 0.479)
$\tau = 0.5$								
intercept	−0.072	(−0.144, −0.001)	−0.006	(−0.221, 0.199)	0.025	(−0.128, 0.147)	−0.008	(−0.253, 0.180)
age	−0.021	(−0.042, −0.001)	−0.045	(−0.091, 0.031)	0.134	(0.076, 0.179)	−0.051	(−0.091, 0.029)
wtkg	−0.005	(−0.024, 0.012)	0.016	(−0.035, 0.091)	−0.175	(−0.210, −0.104)	0.016	(−0.032, 0.095)
race	−0.084	(−0.121, −0.041)	−0.115	(−0.251, 0.017)	−0.051	(−0.182, 0.049)	−0.134	(−0.246, 0.013)
gender	−0.013	(−0.068, 0.042)	−0.046	(−0.212, 0.113)	−0.002	(−0.116, 0.096)	−0.066	(−0.197, 0.108)
history	−0.243	(−0.279, −0.208)	−0.276	(−0.406, −0.148)	−0.217	(−0.285, −0.099)	−0.287	(−0.407, −0.160)
offtrt	−0.384	(−0.422, −0.345)	−0.493	(−0.621, −0.302)	−0.321	(−0.383, −0.183)	−0.500	(−0.623, −0.331)
$CD4_0$	0.509	(0.493, 0.529)	0.531	(0.481, 0.597)	0.523	(0.492, 0.571)	0.517	(0.478, 0.585)
trt	0.372	(0.329, 0.415)	0.366	(0.254, 0.524)	0.231	(0.136, 0.306)	0.385	(0.261, 0.530)
$\tau = 0.75$								
intercept	0.456	(0.397, 0.515)	0.553	(0.319, 0.721)	0.591	(0.506, 0.896)	0.547	(0.328, 0.692)
age	0.037	(0.013, 0.059)	0.005	(−0.051, 0.068)	0.181	(0.119, 0.223)	0.009	(−0.051, 0.069)
wtkg	0.026	(0.004, 0.047)	0.045	(−0.009, 0.102)	−0.163	(−0.207, −0.106)	0.046	(−0.007, 0.106)
race	−0.122	(−0.158, −0.087)	−0.182	(−0.291, −0.057)	−0.095	(−0.131, 0.068)	−0.188	(−0.289, −0.056)
gender	0.018	(−0.029, 0.066)	−0.039	(−0.183, 0.114)	0.041	(−0.205, 0.132)	−0.037	(−0.195, 0.084)
history	−0.207	(−0.242, −0.173)	−0.250	(−0.350, −0.139)	−0.221	(−0.306, −0.109)	−0.255	(−0.348, −0.143)
offtrt	−0.214	(−0.251, −0.177)	−0.403	(−0.559, −0.230)	−0.226	(−0.358, −0.114)	−0.457	(−0.579, −0.266)
$CD4_0$	0.557	(0.536, 0.577)	0.566	(0.514, 0.674)	0.539	(0.489, 0.589)	0.559	(0.505, 0.649)
trt	0.283	(0.235, 0.330)	0.316	(0.181, 0.438)	0.142	(0.118, 0.179)	0.318	(0.203, 0.447)

6. Discussion

In this study, we address the bias in quantile regression estimates by constructing imputation and AIPW estimation equations, with both involving the estimation of conditional means under nonrandom missingness. Many existing methods rely on kernel regression to estimate conditional means. However, nonparametric estimation methods may suffer from the curse of dimensionality when the dimension of the covariates is high. Paik and Larsen [19] proposed using importance resampling to obtain Monte Carlo estimates of conditional means, and Song et al. [21] further applied this method to estimation equations. In this study, we extend these methods to quantile regression and overcome the theoretical and computational challenges caused by the nonsmoothness of the checking function in classical quantile regression by employing convolution smoothing.

Common parameter working models are based on linear regression for observed data. Song et al.'s [21] simulation results showed that model misspecification does not lead to estimation bias. However, their simulation study was based on a regression model that satisfied the Gauss–Markov assumption, with missing response variables following a normal distribution with homoscedasticity concerning the covariates. Misspecification was reflected in the estimation of the mean or location variables. However, the advantages of quantile regression are more evident in situations involving skewness, heavy tails, and heteroscedasticity. In this study, our simulation results under heteroscedasticity showed that imputation estimation based on the assumption of a linear regression working model leads to significant estimation bias, while the AIPW estimation equation can mitigate the impact of model misspecification. We also provide theoretical proof of the consistency of AIPW estimation.

Our simulation results demonstrate that, under the ideal scenario of correctly specified parameter working models, the imputation estimator is more efficient than the IPW and AIPW estimators. The AIPW estimator based on the correctly specified model was found to be more efficient than that based on the misspecified model. Therefore, in practical applications, it is crucial to appropriately specify the parameter working model based on the observed data. Fortunately, the effectiveness of the model specification can be assessed using various methods such as Q-Q plots and histograms. For the observed response conditional distributions that do not conform to the linear regression assumption, a Box–Cox transformation can be applied to approximate a normal parameter working model. If such a parameter working model is difficult to obtain, the AIPW estimator proposed in this study can still provide relatively reliable estimates. This is because the proposed response mechanism model is semiparametric and offers certain flexibility. However, the response model constructed in this study does not consider the interaction effects between covariates X and the response variable Y or the potential nonlinear effects of the response variable Y on the missingness propensity.

Author Contributions: The first two authors contributed equally to this work. J.G.: Methodology, software, validation, data curation, writing; F.L.: visualization, data curation, review; W.K.H.: writing–review and editing; X.Z.: supervision,validation; K.W.: formal analysis; T.Z.: investigation; L.Y.: resources; M.T.: Conceptualization, project administration, funding acquisition, and the corresponding author. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The researchers can download the ACTG175 dataset from the R package "speff2trial".

Abbreviations

The following abbreviations are used in this manuscript:

SIR   Sampling Importance Resampling
IPW   Inverse Probability Weighting
AIPW   Augmented Inverse Probability Weighting
EEI   Imputed Estimation Equation

Appendix A

Appendix A.1. Proof of Lemma 1

$$
\begin{aligned}
E\{\psi_h(\mathbf{Z}_i, Y_i, \boldsymbol{\theta}_0)\} &= E\{\mathbf{Z}_i\{\mathcal{G}_h(\mathbf{Z}_i^\top \boldsymbol{\theta}_0 - Y_i) - \tau\}\} \\
&= E\{\mathbf{Z}_i\{\mathcal{G}(-\varepsilon_i/h) - \tau\}\} \\
&= E\left\{\mathbf{Z}_i\left\{\int_{u < -\varepsilon_i/h} K(u)du - \tau\right\}\right\} \\
&= E\left\{\mathbf{Z}_i E\left\{\int I_{\varepsilon_i < -hu}(\varepsilon_i)K(u)du - \tau \Big| \mathbf{Z}_i\right\}\right\} \\
&= E\left\{\mathbf{Z}_i\left\{\int F(-hu|\mathbf{Z}_i)K(u)du - \tau\right\}\right\}.
\end{aligned}
$$

By assumption (C7), we can utilize a Taylor expansion, thus yielding

$$
\begin{aligned}
&\int F(-hu|\mathbf{Z}_i)K(u)du \\
&= F(0|\mathbf{Z}_i) + \int \sum_{k=1}^{r} F^{(k)}(0|\mathbf{Z}_i)\frac{(-h)^r u^r}{r!}K(u)du + \int F^{(r+1)}(-\tilde{h}u|\mathbf{Z}_i)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du \\
&= F(0|\mathbf{Z}_i) + \int F^{(r)}(0|\mathbf{Z}_i)\frac{(-h)^r u^r}{r!}K(u)du + \int F^{(r+1)}(-\tilde{h}u|\mathbf{Z}_i)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du \\
&= \tau + \int f^{(r-1)}(0|\mathbf{Z}_i)\frac{(-h)^r u^r}{r!}K(u)du + \int f^{(r)}(-\tilde{h}u|\mathbf{Z}_i)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du, \quad \tilde{h} \in [0, h].
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
E\{\mathbf{Z}_i\mathcal{G}_h(\mathbf{Z}_i^\top \boldsymbol{\theta} - Y_i) - \tau\} &= \frac{(-h)^r}{r!}E\{\mathbf{Z}_i f^{(r-1)}(0|\mathbf{Z}_i)\}\int u^r K(u)du \\
&\quad + E[\mathbf{Z}_i f^{(r)}(-\tilde{h}u|\mathbf{Z}_i)u^r K(u)du]O(h^{r+1}).
\end{aligned}
$$

By assumption (C6), we have

$$
\left\|E\left[\mathbf{Z}_i\left\{\int f^{(r)}(-\tilde{h}u|\mathbf{Z}_i)u^r K(u)du\right\}\right]\right\| \leq E\left[\int C(\mathbf{Z})\|\mathbf{Z}\|u^r K(u)du\right] = O(1).
$$

Therefore,

$$
E\{\psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)\} = \frac{(-h)^r}{r!}C_K E[\mathbf{Z}_i f^{(r-1)}(0|\mathbf{Z}_i)] + o(h^r).
$$

Similarly, we have

$$
\begin{aligned}
&E\{m^0_{\psi_h}(\mathbf{Z}_i; \boldsymbol{\theta}_0)\} \\
=&E[E\{\psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)|\mathbf{Z}_i, \delta_i = 0\}] \\
=&E[E\{\mathbf{Z}_i\{\mathcal{G}_h(\mathbf{Z}_i^\top \boldsymbol{\theta}_0 - Y_i) - \tau\}|\mathbf{Z}_i, \delta_i = 0\}] \\
=&E[E\{\mathbf{Z}_i\{\mathcal{G}_h(-\varepsilon_i/h) - \tau\}|\mathbf{Z}_i, \delta_i = 0\}] \\
=&E\left[E\left\{\mathbf{Z}_i\left\{\int_{u<-\varepsilon_i/h} K(u)du - \tau\right\}|\mathbf{Z}_i, \delta_i = 0\right\}\right] \\
=&E\left[E\left\{\mathbf{Z}_i\left\{\int I_{\varepsilon_i<-hu}(\varepsilon_i)K(u)du - \tau\right\}|\mathbf{Z}_i, \delta_i = 0\right\}\right] \\
=&E\left[E\left\{\mathbf{Z}_i\left\{\int F(-hu|\mathbf{Z}_i, \delta_i = 0)K(u)du - \tau\right\}|\mathbf{Z}_i, \delta_i = 0\right\}\right].
\end{aligned}
$$

Based on the assumptions and the Taylor expansion, we have

$$
\begin{aligned}
&\int F(-hu|\mathbf{Z}_i, \delta_i = 0)K(u)du \\
=&F(0|\mathbf{Z}_i, \delta_i = 0) + \int \sum_{k=1}^{r} F^{(k)}(0|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^r u^r}{r!}K(u)du \\
&+ \int F^{(r+1)}(-\tilde{h}u|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du \\
=&F(0|\mathbf{Z}_i, \delta_i = 0) + \int F^{(r)}(0|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^r u^r}{r!}K(u)du \\
&+ \int F^{(r+1)}(-\tilde{h}u|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du \\
=&F(0|\mathbf{Z}_i, \delta_i = 0) + \int f^{(r-1)}(0|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^r u^r}{r!}K(u)du \\
&+ \int f^{(r)}(-\tilde{h}u|\mathbf{Z}_i, \delta_i = 0)\frac{(-h)^{r+1}u^{r+1}}{(r+1)!}K(u)du, \quad \tilde{h} \in [0, h].
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&E[E\{\mathbf{Z}_i\{F(0|\mathbf{Z}_i, \delta_i = 0) - \tau\}|\mathbf{Z}_i, \delta_i = 0\}] \\
=&E\left[\mathbf{Z}_i E\{I(Y_i < \mathbf{Z}_i^\top \boldsymbol{\theta}_0) - \tau\}|\mathbf{Z}_i, \delta_i = 0\right] \\
=&E\left[E\{\mathbf{Z}_i\{I(Y_i < \mathbf{Z}_i^\top \boldsymbol{\theta}_0) - \tau\}|\mathbf{Z}_i, \delta_i = 0\}\right] \\
=&E\{m^0_\psi(\mathbf{Z}_i; \boldsymbol{\theta}_0)\},
\end{aligned}
$$

which implies

$$
\begin{aligned}
&E\{m^0_{\psi_h}(\mathbf{Z}_i; \boldsymbol{\theta}_0)\} \\
=&E[E\{\psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)|\mathbf{Z}_i, \delta_i = 0\}] \\
=&E\{m^0_\psi(\mathbf{X}_i; \boldsymbol{\theta}_0)\} + \frac{(-h)^r}{r!}E\{\mathbf{Z}_i f^{(r-1)}(0|\mathbf{Z}_i, \delta_i = 0)\}\int u^r K(u)du \\
&+ E[\mathbf{Z}_i\{f^{(r)}(-\tilde{h}u|\mathbf{Z}_i, \delta_i = 0)u^r K(u)du\}]O(h^{r+1}).
\end{aligned}
$$

Under the assumption conditions, we have

$$
\left\|E\left[\mathbf{Z}_i\left\{\int f^{(r)}(-\tilde{h}u|\mathbf{Z}_i, \delta_i = 0)u^r K(u)du\right\}\right]\right\| \leq E\left[\int C(\mathbf{Z})\|\mathbf{Z}\|u^r K(u)du\right] = O(1),
$$

which implies

$$
E\{m^0_{\psi_h}(\mathbf{Z}_i; \boldsymbol{\theta}_0)\} = E\{m^0_\psi(\mathbf{Z}_i; \boldsymbol{\theta}_0)\} + \frac{(-h)^r}{r!}C_K E[\mathbf{Z}_i f^{(r-1)}(0|\mathbf{Z}_i, \delta_i = 0)] + o(h^r).
$$

Appendix A.2. Proof of Lemma 2

To prove (1), we can perform a simple calculation. We have

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{hi}^{(1)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0) + (1 - \delta_i) \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0) + (1 - \delta_i) m_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - \delta_i) \{ \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) - m_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) \} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - \delta_i) \{ \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) \} \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0) + (1 - \delta_i) m_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) \\
& + \frac{1}{\sqrt{n}} (1 - \delta_i) \left\{ \frac{\partial \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}^{*}, \gamma^{*})}{\partial \boldsymbol{\beta}^{\top}} \right\}^{\top} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
& + \frac{1}{\sqrt{n}} (1 - \delta_i) \left\{ \frac{\partial \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}^{*}, \gamma^{*})}{\partial \gamma} \right\}^{\top} (\hat{\gamma} - \gamma_0) + o_p(1) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{i1} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{i2} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{i3} + o_p(1).
\end{aligned}
$$

Based on the fact that $E\{\psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)\} = O(h^r)$ and

$$
E\{m_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0)\} = E\{m_{\psi}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0)\} + O(h^r), \tag{A1}
$$

we have

$$
EI_{i1} = E\{\psi_h(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)\} = O(h^r).
$$

Additionally, we have

$$
EI_{i1}^{2} = E\left[\mathbf{Z}_i \mathbf{Z}_i^{\top} \{ \delta_i \mathcal{G}_h(\mathbf{Z}_i^{\top} \boldsymbol{\theta}_0 - Y_i) + (1 - \delta_i) E\{ \mathcal{G}_h(\mathbf{Z}_i^{\top} \boldsymbol{\theta}_0 - Y_i) | \mathbf{Z}_i, \delta_i = 0 \} \}^{\otimes 2} \right],
$$

According to the assumptions, as $n \to \infty$,

$$
\begin{aligned}
\lim_{nh^{2r} \to 0} EI_{i1}^{2} =& E\left[\left\{ \delta_i \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0) + (1 - \delta_i) m_{\psi}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0) \right\}^{\otimes 2} \right] \\
=& E\left[\pi(\mathbf{X}_i, Y_i) \psi(\mathbf{Z}_i, Y_i; \boldsymbol{\theta}_0)^{\otimes 2} + (1 - \pi(\mathbf{X}_i, Y_i)) m_{\psi}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0)^{\otimes 2} \right] \\
:=& A_1,
\end{aligned}
$$

thus yielding

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{i1} \xrightarrow{d} N(0, A_1).
$$

For I_{i2}, we have

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{i2} =& \frac{1}{n} \sum_{i=1}^{n} (1 - \delta_i) \left\{ \frac{\partial \hat{m}_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}^{*}, \gamma^{*})}{\partial \boldsymbol{\beta}^{\top}} \right\}^{\top} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
=& \frac{1}{n} \sum_{i=1}^{n} (1 - \delta_i) \left\{ \frac{\partial m_{\psi_h}^{0}(\mathbf{Z}_i; \boldsymbol{\theta}_0, \boldsymbol{\beta}^{*}, \gamma^{*})}{\partial \boldsymbol{\beta}^{\top}} \right\}^{\top} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1).
\end{aligned}
$$

where

$$\lim_{nh^{2r}\to 0} E\left\{ (1-\delta_i)\left\{ \frac{\partial m_{\psi_h}^0(\mathbf{Z}_i;\boldsymbol{\theta}_0,\boldsymbol{\beta}^*,\gamma^*)}{\partial \boldsymbol{\beta}^\top} \right\}^\top \right\}$$
$$= E\{(1-\delta_i)\mathrm{Cov}_0(\psi(\mathbf{Z}_i,Y_i;\boldsymbol{\theta}_0),s(\mathbf{Z}_i,Y_i;\boldsymbol{\beta}_0)|\mathbf{Z}_i)\} + o_p(1).$$

Let $\mathbf{H}_1 = E\{(1-\delta_i)\mathrm{Cov}_0(\psi(\mathbf{Z}_i,Y_i;\boldsymbol{\theta}_0),s(\mathbf{Z}_i,Y_i;\boldsymbol{\beta}_0)|\mathbf{Z}_i)\}$. By the assumption, we have $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$ and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Sigma})$. Therefore, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i2} \xrightarrow{d} N(\mathbf{0},\mathbf{H}_1\boldsymbol{\Sigma}\mathbf{H}_1^\top).$$

Similarly, let $\mathbf{H}_2 = E\{(1-\delta_i)\mathrm{Cov}_0(\psi(\mathbf{Z}_i,Y_i;\boldsymbol{\theta}_0),Y|\mathbf{Z}_i)\}$. According to Shao [5], we have $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$ and $\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0,\sigma^2)$. Thus, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i3} \xrightarrow{d} N(\mathbf{0},\sigma^2\mathbf{H}_2^{\otimes 2}).$$

It can be shown that $E\{I_{i1} + I_{i2}\} = o_p(1)$. We have

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i1}\right) \cdot \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i2}\right) = \frac{1}{n}\sum_{i=1}^n I_{i1}I_{i2} + \frac{1}{n}\sum_{i\neq j}^n\sum_{j=1}^n I_{i1}I_{j2}.$$

where

$$\frac{1}{n}\sum_{i=1}^n I_{i1}I_{i2} = \frac{1}{n}\sum_{i=1}^n \psi_h^{(l)}(\mathbf{Z}_i,Y_i,\delta_i;\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)\left\{ \frac{\partial \hat{m}_{\psi_h}^0(\mathbf{Z}_i;\boldsymbol{\theta}_0,\boldsymbol{\beta}^*,\gamma^*)}{\partial \boldsymbol{\beta}^\top} \right\}^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$= o_p(1),$$

$$\frac{1}{n}\sum_{i\neq j}^n\sum_{j=1}^n I_{i1}I_{j2} = \frac{1}{n}\sum_{i\neq j}\sum_{j=1}^n \psi_{hi}^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)\left\{ \frac{\partial m_{\psi_h}^0(\mathbf{Z}_j;\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)}{\partial \boldsymbol{\beta}^\top} \right\}^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$+ o_p(n^{-1/2}).$$

For $i \neq j$, $\psi_{hi}^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)$ and $\left\{ \frac{\partial m_{\psi_h}^0(\mathbf{Z}_j;\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)}{\partial \boldsymbol{\beta}^\top} \right\}$ are independent; therefore,

$$E\left[\psi_{hi}^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0) \right] = O(h^r).$$
$$E\left[\left\{ \frac{\partial m_{\psi_h}^0(\mathbf{Z}_j;\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)}{\partial \boldsymbol{\beta}^\top} \right\}^\top \right](\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(n^{-1/2}).$$

By the assumption, we have $\frac{1}{n}\sum_{i\neq j}^n\sum_{j=1}^n I_{i1}I_{j2} = (n-1)O(h^r)O_p(n^{-1/2}) = O_p(n^{1/2}h^r) = o_p(1)$. Hence, we can conclude that $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i1}\right) \cdot \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i2}\right) = o_p(1)$, which implies

$$\mathrm{Cov}(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i1}, \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i2}) = o(1).$$

Similarly, we can show that $\mathrm{Cov}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i1}, \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i3}\right) = o(1)$. Consequently,

$$\mathrm{Cov}(\frac{1}{\sqrt{n}}\sum_{i=1}^n I_{i1}, \frac{1}{\sqrt{n}}\sum_{i=1}^n (I_{i2} + I_{i3})) = o(1).$$

To establish the asymptotic properties of $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_{i2}\right)\cdot\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_{i3}\right)$, we employ Taylor expansions, thus yielding the following:

$$\sqrt{n}(\hat{\gamma}-\gamma_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\gamma_0)+o_p(1),\quad\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_i(\boldsymbol{\beta}_0)+o_p(1).$$

Consequently, as $n\to\infty$, we have

$$\frac{1}{\sqrt{n}}(I_{i2}+I_{i3})\overset{d}{\to}N(0,\boldsymbol{B}_1),$$

where $\boldsymbol{B}_1:=\mathrm{Var}(\boldsymbol{H}_1\boldsymbol{\Psi}(\boldsymbol{\beta}_0)+\boldsymbol{H}_2\Phi(\gamma_0))$. Furthermore, we obtain

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{hi}^{(1)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})\overset{d}{\to}N(0,\boldsymbol{T}_1),$$

where $\boldsymbol{T}_1=\boldsymbol{A}_1+\boldsymbol{B}_1$.

To investigate the asymptotic properties of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{hi}^{(2)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{hi}^{(2)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{hi}^{(2)}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\pi(\boldsymbol{X}_i,Y_i;\hat{g}_{\gamma_0},\gamma_0)}-\frac{\delta_i}{\pi(\boldsymbol{X}_i,Y_i)}\right\}\left\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right\}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\pi(\boldsymbol{X}_i,Y_i;\hat{g}_{\hat{\gamma}},\hat{\gamma})}-\frac{\delta_i}{\pi(\boldsymbol{X}_i,Y_i;\hat{g}_{\gamma_0},\gamma_0)}\right\}\left\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right\}+o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i1}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i2}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i3}+o_p(1).$$

Similar to the previous proof, as $n\to\infty$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i1}\overset{d}{\to}N(0,\boldsymbol{A}_2),$$

where $\boldsymbol{A}_2=E\left\{\pi(\boldsymbol{X}_i,Y_i)^{-1}(\psi(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0))^{\otimes 2}\right\}+E\left\{m_{\psi}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)^{\otimes 2}\right\}$.

To analyze the asymptotic behavior of $\frac{1}{\sqrt{n}}L_{i2}$, we have

$$\frac{1}{\sqrt{n}}L_{i2}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1-\frac{\delta_i}{\pi(\boldsymbol{X}_i,Y_i)}\right\}\left\{E\left\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)|\boldsymbol{X}_i,\delta_i=0\right\}\right\}+o_p(1)$$
$$=o_p(1).$$

According to the analysis, we can conclude that $\frac{1}{\sqrt{n}}L_{i2}$ converges to zero in probability, i.e., $\frac{1}{\sqrt{n}}L_{i2}=o_p(1)$.

To establish the asymptotic properties of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i3}$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i3}$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\frac{\partial\pi^{-1}(\boldsymbol{X}_i,Y_i;\hat{g}_{\gamma},\gamma)}{\partial\gamma}\Big\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\Big\}(\hat{\gamma}-\gamma_0)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\Big\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\Big\}\exp(-\hat{g}_{\gamma_0}(\boldsymbol{X}_i)+\gamma_0 Y_i)\{Y_i-\hat{m}_Y^0(\boldsymbol{X}_i;\gamma_0)\}(\hat{\gamma}-\gamma_0)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i O(\boldsymbol{X}_i,Y_i)\Big\{\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)-m_{\psi_h}^0(\boldsymbol{Z}_i,\boldsymbol{\theta}_0)\Big\}\{Y_i-m_Y^0(\boldsymbol{X}_i;\gamma_0)\}(\hat{\gamma}-\gamma_0)+o_p(1)$$

$$=:\boldsymbol{H}_3\sqrt{n}(\hat{\gamma}-\gamma_0)+o_p(1),$$

where $\boldsymbol{H}_3=E\Big\{(1-\delta)(Y-m_Y^0(\boldsymbol{X}))(\psi(\boldsymbol{Z},Y;\boldsymbol{\theta}_0)-m_{\psi}^0(\boldsymbol{Z};\boldsymbol{\theta}_0))\Big\}$.

According to the Slutzky theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i3}\xrightarrow{d}N(0,\boldsymbol{B}_2),$$

where $\boldsymbol{B}_2=\mathrm{Var}(\boldsymbol{H}_3\Phi(\gamma_0))$. Similarly to the previous proof, it can be shown that

$$\mathrm{Cov}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i1},\frac{1}{\sqrt{n}}\sum_{i=1}^{n}L_{i3})=o_p(1),$$

which implies that, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\psi}_{hi}^{(2)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})\xrightarrow{d}N(0,\boldsymbol{T}_2),$$

where $\boldsymbol{T}_2=\boldsymbol{A}_2+\boldsymbol{B}_2$.

To prove (2), we first establish the asymptotic property of $\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_{hi}^{(1)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})^{\otimes 2}$. By the law of large numbers and the fact that $\hat{\gamma}-\gamma_0=o_p(1)$ and $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=o_p(1)$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_{hi}^{(1)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})^{\otimes 2}=\frac{1}{n}\sum_{i=1}^{n}\psi_{hi}^{(1)}(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0,\gamma_0)^{\otimes 2}+o_p(1).$$

As $n\to\infty$, under the assumption, we have

$$\lim_{nh^{2r}\to 0}E\Big\{\Big[\delta_i\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)+(1-\delta_i)m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\Big]^{\otimes 2}\Big\}$$

$$=\lim_{nh^{2r}\to 0}E\Big[\delta_i\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)^{\otimes 2}+(1-\delta_i)m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)^{\otimes 2}\Big]$$

$$\quad+\lim_{nh^{2r}\to 0}E\Big[2\delta_i(1-\delta_i)\psi_h(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)m_{\psi_h}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\Big]$$

$$=E\Big[\delta_i\psi(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)^{\otimes 2}+(1-\delta_i)m_{\psi}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)^{\otimes 2}+2\delta_i(1-\delta_i)\psi(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)m_{\psi}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\Big]$$

$$=E\Big[\delta_i\psi(\boldsymbol{Z}_i,Y_i;\boldsymbol{\theta}_0)^{\otimes 2}+(1-\delta_i)m_{\psi}^0(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)^{\otimes 2}\Big]$$

$$:=\boldsymbol{A}_1.$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_{hi}^{(1)}(\boldsymbol{\theta}_0,\hat{\boldsymbol{\beta}},\hat{\gamma})^{\otimes 2}\xrightarrow{p}\boldsymbol{A}_1.$$

Similarly, for $\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(2)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2}$, we have

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(2)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} = \frac{1}{n} \sum_{i=1}^{n} \psi_{hi}^{(2)}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0)^{\otimes 2} + o_p(1).$$

As $n \to \infty$, under the assumption, we have

$$\lim_{nh^{2r} \to 0} E\left\{ \left[\frac{\delta_i}{\pi(\boldsymbol{X}_i, Y_i)} \{\psi_h(\boldsymbol{Z}_i, Y_i; \boldsymbol{\theta}_0) - m_{\psi_h}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\} + m_{\psi_h}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right]^{\otimes 2} \right\}$$

$$= \lim_{nh^{2r} \to 0} E\left\{ \left[\frac{\delta_i}{\pi(\boldsymbol{X}_i, Y_i)^2} \{\psi_h(\boldsymbol{Z}_i, Y_i; \boldsymbol{\theta}_0) - m_{\psi_h}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\}^{\otimes 2} + \{m_{\psi_h}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\}^{\otimes 2} \right] \right\}$$

$$= E\left\{ \pi(\boldsymbol{X}_i, Y_i)^{-1} \{\psi(\boldsymbol{Z}_i, Y_i; \boldsymbol{\theta}_0) - m_{\psi}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\}^{\otimes 2} + \{m_{\psi}^0(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\}^{\otimes 2} \right\}$$

$$= A_2.$$

Therefore, we have

$$\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(2)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \xrightarrow{p} A_2.$$

Next, we prove (3). Note that, for $l = 1, 2$, we have

$$E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \psi_{hi}^{(l)}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \gamma_0) \right\} = E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \psi_h(\boldsymbol{Z}_i, Y_i; \boldsymbol{\theta}_0) \right\}$$

$$= E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{Z}_i G_h(\boldsymbol{Z}_i^\top \boldsymbol{\theta} - Y_i) \right\} = E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{Z}_i \int I(Y_i < \boldsymbol{Z}_i^\top \boldsymbol{\theta}_0 - uh) K(u) du \right\}$$

$$= E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{Z}_i \int F_Y(\boldsymbol{Z}_i^\top \boldsymbol{\theta}_0 - uh) K(u) du \right\} = E\left\{ \boldsymbol{Z}_i \int \frac{\partial}{\partial \boldsymbol{\theta}^\top} F_Y(\boldsymbol{Z}_i^\top \boldsymbol{\theta}_0 - uh) K(u) du \right\}$$

$$= E\left\{ \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \int f_Y(\boldsymbol{Z}_i^\top \boldsymbol{\theta}_0 - uh) K(u) du \right\} = E\left\{ \boldsymbol{Z}_i \boldsymbol{Z}_i^\top f_Y(\boldsymbol{Z}_i^\top \boldsymbol{\theta}_0 | \boldsymbol{Z}_i) \right\} + o_p(1)$$

$$= E\left\{ \boldsymbol{Z}_i \boldsymbol{Z}_i^\top f(0 | \boldsymbol{Z}_i) \right\} + o_p(1) := \boldsymbol{\Gamma} + o_p(1).$$

By the law of large numbers, as $n \to \infty$, we have $\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{\partial \boldsymbol{\theta}^\top} \xrightarrow{p} \boldsymbol{\Gamma}$.

Finally, we demonstrate (4). From

$$\frac{n^{-1}(\max_i \|\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\|)^2}{n^{-1} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2}} \to 0,$$

it can be easily shown that, for $l = 1, 2$, $\max_i \|\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\| = o_p(n^{1/2})$.

Proof of Theorem 1. By applying the Lagrange multiplier method, we obtain the empirical log-likelihood ratio function with respect to the parameter vector $\boldsymbol{\theta}$:

$$\hat{R}^{(l)}(\boldsymbol{\theta}) = 2 \sum_{i=1}^{n} \log\{1 + \boldsymbol{\lambda}^\top \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})\},$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta})$ is the solution to the following equation:

$$g(\boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{1 + \boldsymbol{\lambda}^\top \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})}.$$

In other words, $\hat{\boldsymbol{\theta}}$ simultaneously satisfies the following two equations:

$$T_{1n}^{(l)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{1 + \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})},$$

$$T_{2n}^{(l)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\{\partial \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma}) / \partial \boldsymbol{\theta}^{\top}\} \boldsymbol{\lambda}}{1 + \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}, \hat{\gamma})}.$$

Note that $T_{1n}^{(l)}(\hat{\boldsymbol{\theta}}, 0) = n^{-1} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\hat{\boldsymbol{\theta}})$, and $T_{2n}^{(l)}(\hat{\boldsymbol{\theta}}, 0) = 0$. Under the assumption conditions, according to Lemma A.1 in Newey and Smith [30] and Theorem 1(a) in Leng and Tang Leng [31], it can be shown that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\theta 0$. By Taylor expanding $T_{1n}^{(l)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ and $T_{2n}^{(l)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ around $(\boldsymbol{\theta}_0, 0)$, we have

$$0 = T_{1n}^{(l)}(\boldsymbol{\theta}_0, 0) + \frac{\partial T_{1n}^{(l)}(\boldsymbol{\theta}_0, 0)}{\partial \boldsymbol{\theta}^{\top}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{\partial T_{1n}^{(l)}(\boldsymbol{\theta}_0, 0)}{\partial \boldsymbol{\lambda}} \boldsymbol{\lambda} + o_p(u_n),$$

$$0 = T_{2n}^{(l)}(\boldsymbol{\theta}_0, 0) + \frac{\partial T_{2n}^{(l)}(\boldsymbol{\theta}_0, 0)}{\partial \boldsymbol{\theta}^{\top}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{\partial T_{2n}^{(l)}(\boldsymbol{\theta}_0, 0)}{\partial \boldsymbol{\lambda}} \boldsymbol{\lambda} + o_p(u_n),$$

where $u_n = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + |\boldsymbol{\lambda}\|$.

The above equations can be rewritten as follows:

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{n} \sum_{i=1}^{n} \hat{\boldsymbol{\psi}}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} & \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{\partial \boldsymbol{\theta}^{\top}} \\ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{\partial \boldsymbol{\theta}^{\top}} & 0 \end{pmatrix}^{-1}$$
$$\begin{pmatrix} -\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) + o_p(u_n) \\ o_p(u_n) \end{pmatrix}.$$

Based on the results of Lemma 2, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^{(l)} - \boldsymbol{\theta}_0) = \left\{ -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{\partial \boldsymbol{\theta}^{\top}} \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) + o_p(1).$$

Therefore, for $l = 1, 2$ and as $n \to \infty$, we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^{(l)} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}^{-1} \boldsymbol{T}_l \boldsymbol{\Gamma}^{-\top}).$$

□

Proof of Theorem 2. First, we note that $\|\boldsymbol{\lambda}\| = O_p(n^{-1/2})$. Let $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}_0) = \rho \boldsymbol{u}$, where $\boldsymbol{u} = \boldsymbol{\lambda} / \|\boldsymbol{\lambda}\|$ and $\|\boldsymbol{u}\| = 1$. We have

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{1 + \boldsymbol{\lambda}(\boldsymbol{\theta})^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}{1 + \rho \boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{u} \rho}{1 + \rho \boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})},$$

By multiplying both sides of the equation by $\boldsymbol{u}^{\top}$, we obtain

$$
\begin{aligned}
&\left\| \boldsymbol{u}^{\top} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \right\| \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{u} \rho}{1 + \rho \boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})} \\
&\geq \frac{1}{1 + \rho \max_i |\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})|} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{u} \rho.
\end{aligned}
$$

Thus, we can conclude that

$$
\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{u} \rho \leq \left\| \boldsymbol{u}^{\top} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \right\| \{1 + \rho \max_i \|\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\|\}.
$$

Based on Lemma 2, we have

$$
\rho \boldsymbol{u}^{\top} A_l \boldsymbol{u} + o_p(1) \leq O_p(n^{-1/2}) 1 + \rho o_p(n^{1/2}).
$$

Consequently, it follows that $\rho = O_p(n^{-1/2})$. Furthermore, we can observe that

$$
\max_i \|\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\| \leq \|\boldsymbol{\lambda}\| \max_i \|\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\| = O_p(n^{-1/2}) o_p(n^{1/2}) = o_p(1).
$$

By expanding the function $g(\boldsymbol{\lambda})$, we obtain

$$
\begin{aligned}
0 = g(\boldsymbol{\lambda}) &= \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \left[1 - \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) + \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^2}{(1 + \eta_i)^3} \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \boldsymbol{\lambda} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^2}{(1 + \eta_i)^3},
\end{aligned}
$$

where $\eta_i \in (0, \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}))$. From the fact that $\max_i \|\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\| = o_p(1)$, it follows that $|\xi_i| = o_p(1)$.

Note that

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^2}{(1 + \eta_i)^3} \right\| \\
&\leq \frac{\max_i \|\hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})\|}{1 - \max_i |\xi_i|} \left\| \boldsymbol{\lambda}^{\top} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \boldsymbol{\lambda} \right\| \\
&= o_p(n^{1/2}) O_p(n^{-1}) = o_p(n^{-1/2}).
\end{aligned}
$$

Therefore,

$$
\boldsymbol{\lambda} = \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \right\} + \boldsymbol{\zeta},
$$

where $\|\boldsymbol{\zeta}\| = o_p(n^{-1/2})$.

By expanding $\hat{R}^{(l)}(\boldsymbol{\theta}_0)$ around $\boldsymbol{\theta}_0$ using a Taylor series, we obtain

$$
\begin{aligned}
\hat{R}^{(l)}(\boldsymbol{\theta}_0) =& 2 \sum_{i=1}^{n} \left[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \frac{1}{2} [\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^2 + \frac{1}{3} \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^3}{(1 + \xi_i)^3} \right] \\
=& 2 \boldsymbol{\lambda}^{\top} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \sum_{i=1}^{n} \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{\lambda} \\
& + \frac{2}{3} \sum_{i=1}^{n} \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^3}{(1 + \xi_i)^3}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\left| \sum_{i=1}^{n} \frac{[\boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})]^3}{(1 + \xi_i)^3} \right| \leq& \frac{\max_i \| \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \|}{1 - \max_i |\xi_i|} \| \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{\lambda} \| \\
=& o_p(1) n O_p(n^{-1}) = o_p(1).
\end{aligned}
$$

Therefore, we obtain

$$
\hat{R}^{(l)}(\boldsymbol{\theta}_0) = 2 \boldsymbol{\lambda}^{\top} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) - \sum_{i=1}^{n} \sum_{i=1}^{n} \boldsymbol{\lambda}^{\top} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \boldsymbol{\lambda} + o_p(1).
$$

By combining the previous results, we can express

$$
\begin{aligned}
\hat{R}^{(l)}(\boldsymbol{\theta}_0) =& n \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \right\}^{\top} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \right\}^{-1} \\
& - n \boldsymbol{\zeta}^{\top} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \boldsymbol{\zeta} + o_p(1).
\end{aligned}
$$

Here, $n \boldsymbol{\zeta}^{\top} \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \boldsymbol{\zeta} = n o_p(n^{-1/2}) o_p(n^{-1/2}) = o_p(1)$. Thus,

$$
\begin{aligned}
\hat{R}^{(l)}(\boldsymbol{\theta}_0) =& n \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma}) \right\}^{\top} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\otimes 2} \right\}^{-1} \\
& \cdot \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\psi}_{hi}^{(l)}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\beta}}, \hat{\gamma})^{\top} \right\} \\
& + o_p(1).
\end{aligned}
$$

Based on the results (1) and (2) of Lemma 2, it can be easily demonstrated that, as n tends to infinity, the asymptotic distribution of $\hat{R}^{(l)}(\boldsymbol{\theta}_0)$ follows a linear combination of independent chi-squared random variables:

$$
\hat{R}^{(l)}(\boldsymbol{\theta}_0) \xrightarrow{d} r_1^{(l)} \chi_{1 \cdot 1}^2 + r_2^{(l)} \chi_{1 \cdot 2}^2 + \cdots + r_q^{(l)} \chi_{1 \cdot q}^2,
$$

where $r_i^{(l)}$ represents the eigenvalues of $A_l^{-1} T_l$. Here, $\chi_{1 \cdot 1}^2, \chi_{1 \cdot 2}^2, \ldots, \chi_{1 \cdot q}^2$ denote q independent standard chi-squared distributed random vectors. This completes the proof of the theorem. □

References

1. Robins, J.M.; Ritov, Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi–parametric models. *Stat. Med.* **1997**, *16*, 285–319. [CrossRef]
2. Wang, S.; Shao, J.; Kim, J.K. An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Stat. Sin.* **2014**, *24*, 1097–1116. [CrossRef]
3. Kenward, M.G. Selection models for repeated measurements with non–random dropout: An illustration of sensitivity. *Stat. Med.* **1998**, *17*, 2723–2732. [CrossRef]
4. Kim, J.K.; Yu, C.L. A semiparametric estimation of mean functionals with nonignorable missing data. *J. Am. Stat. Assoc.* **2011**, *106*, 157–165. [CrossRef]
5. Shao, J.; Wang, L. Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika* **2016**, *103*, 175–187. [CrossRef]
6. Kim, J.K.; Shao, J. *Statistical Methods for Handling Incomplete Data*; CRC Press: New York, NY, USA, 2022.
7. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
8. Koenker, R.; Bassett, G. Tests of linear hypotheses and L1 estimation. *Econometrica* **1982**, *50*, 1577–1583. [CrossRef]
9. Horowitz, J.L. Bootstrap methods for median regression models. *Econometrica* **1998**, *66*, 1327–1351. [CrossRef]
10. Whang, Y.J. Bootstrap methods for median regression models. *Econ. Theory* **2006**, *22*, 173–205.
11. Luo, S.H.; Mei, C.L.; Zhang, C.Y. Smoothed empirical likelihood for quantile regression models with response data missing at random. *Adv. Stat. Anal.* **2017**, *101*, 95–116. [CrossRef]
12. Zhang, T.; Wang, L. Smoothed empirical likelihood inference and variable selection for quantile regression with nonignorable missing response. *Comput. Stat. Data. Anal.* **2020**, *144*, 106888. [CrossRef]
13. Niu, C.; Guo, X.; Xu, W.; Zhu, L. Empirical likelihood inference in linear regression with nonignorable missing response. *Comput. Stat. Data. Anal.* **2014**, *79*, 91–112. [CrossRef]
14. Bindele, H.F.; Zhao, Y.C. Rank-based estimating equation with non-ignorable missing responses via empirical likelihood. *Stat. Sin.* **2018**, *28*, 1787–1820. [CrossRef]
15. Chen, X.R.; Wan, A.T.K.; Zhou, Y. Efficient quantile regression analysis with missing observations. *J. Am. Stat. Assoc.* **2015**, *110*, 723–741. [CrossRef]
16. Tang, N.S.; Zhao, P.Y.; Zhu, H.T. Efficient quantile regression analysis with missing observations. *Stat. Sin.* **2014**, *24*, 723–747. [PubMed]
17. Kim, J.K. Parametric fractional imputation for missing data analysis. *Biometrika* **2011**, *98*, 119–132. [CrossRef]
18. Riddles, M.K.; Kim, J.K.; Im, J. A propensity-score-adjustment method for nonignorable nonresponse. *J. Surv. Stat. Methodol.* **2016**, *4*, 215–245. [CrossRef]
19. Paik, M.; Larsen, M.D. Handling nonignorable nonresponse with respondent modeling and the SIR algorithm. *J. Stat. Plan. Inference* **2014**, *145*, 179–189. [CrossRef]
20. Wang, X.L.; Song, Y.Q.; Lin, L. Handling estimating equation with nonignorably missing data based on SIR algorithm. *J. Comput. Appl. Math.* **2017**, *326*, 62–70. [CrossRef]
21. Song, Y.Q.; Zhu, Y.J.; Wang, X.L.; Lin, L. Robust inference for estimating equations with nonignorably missing data based on SIR algorithm. *J. Stat. Comput. Simul.* **2019**, *89*, 3196–3212. [CrossRef]
22. Newey, W.K.; McFadden, D. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*; Engle, R.F., McFadden, D., Eds.; Elsevier: Amsterdam, The Netherlands, 1994; pp. 2111–2245.
23. Van der Vaart, A.W. Semiparametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1999)*; Bernard, P., Ed.; Springer: Berlin, Germany, 2002; pp. 331–457.
24. Morikawa, K.; Kim, J.K.; Kano, Y. Semiparametric maximum likelihood estimation with data missing not at random. *Can. J. Stat.* **2017**, *45*, 393–409. [CrossRef]
25. Morikawa, K.; Kim, J.K. Semiparametric optimal estimation with nonignorable nonresponse data. *Ann. Stat.* **2021**, *49*, 2991–3014. [CrossRef]
26. Zhao, P.; Wang, L.; Shao, J. Empirical likelihood and Wilks phenomenon for data with nonignorable missing values. *Scan. J. Stat.* **2019**, *46*, 1003–1024. [CrossRef]
27. Hammer, S.M.; Katzenstein, D.A.; Hughes, M.D.; Gundacker, H.; Schooley, R.T.; Haubrich, R.H.; Henry, W.K.; Lederman, M.M.; Phair, J.P.; Niu, M.; et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *N. Engl. J. Med.* **1996**, *335*, 1081–1090. [CrossRef] [PubMed]
28. Davidian, M.; Tsiatis, A.A.; Leon, S. Semiparametric estimation of treatment effect in a pretest–posttest study with missing data. *Stat. Sci.* **2005**, *20*, 261–301. [CrossRef] [PubMed]
29. Zhang, M.; Tsiatis, A.A.; Davidian, M. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics* **2008**, *64*, 707–715. [CrossRef]

30. Newey, W.K.; Smith, R.J. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **2004**, *72*, 219–255. [CrossRef]
31. Leng, C.L.; Tang, C.Y. Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika* **2012**, *99*, 703–716. [CrossRef]